


```
QY 710 CTTTCGGCGGACGACCAACATCGGATCGTGTGTACAGCGGCGGACGTACTCTTG 769
Db 554 CTTTCGGCGGACGACCAACATCGGATCGTGTGTACAGCGGCGGACGTACTCTATATG 495
QY 770 ACGACGTAACCAACCGCCCAAGGTTCCGCGAGACGCTTCCCACTTAGCGAAATCT 829
Db 494 ACGGCGGCAAGCTCGCGCCCGGTCTCTTTCGAGCTGCTCGGCAACTGAAAGCGTGA 435
QY 830 CTCGCCCTGCTACCTACCTGTCCTGCGGAAAGCTGGAGAAATTAGTGGGTGCCCTTGAGC 889
Db 434 TCGCGACGCTGCTATTTTCACGTGCCGCGGCTTCGACATGATCGCGCATTTGCGCG 375
QY 890 GAGACAGTACCTTCGCGGAAAGCTTCTTCGCTCGCATGAAAGCTTCTTCGCGCGCG 949
Db 374 GTACCAAGAGGTTTCGCGCGCGCTCTTCACGAGGAGGAAATTCGCTTTCAGCGCGCG 315
QY 950 CTGGGTTGTCGCAAGGATCTGGGATCTGTTGACCGGGTCTGTAACAGCACTGTGTG 1009
Db 314 CACGCGTGGCGGAAACCTGTGGGACGCGCTCGAAACAATTGCTCGTGAACCGTGGCC 255
QY 1010 AGGGCAATTCGATGATGCGGGGTCTGGGCAATGACGAGACTGCTCTTCTGCACTTTA 1069
Db 254 GCCCGCTGCGATGATTCGCGCTGGGGCTCGACGGAACCTCGCGCTGGCGACT 195
QY 1070 CCAACCGGACCGCTGTCGATGCTGTGTACATTTGGGCTGCCAGCGCTGCTGGAGGTCA 1129
Db 194 GCCATTTCTTCGCGGAGCGCTCGGGCAATTCGCGCTGCCGATACCGGTAACGAGTGA 135
QY 1130 AGCTCGTTCGCGTGTGATGGAATTTGGAAGGGGCTTTCATGTTCGCACTGATGAGCG 1189
Db 134 AGCTCGTTCATCTCGGGGACACAGCTGGAATGCGCGCGCGCCCAAGTCAACGCGCG 75
QY 1190 GACTACTGGCGTGTCTCGTAACAATGCCAAGGTTTCGACGAGGAAGCTATTACGCT 1249
Db 74 GTTACTGGAAGGCGCCGAGCTGACAGGAGGTTTGAAGAGGAGCTTTATCTCA 15
QY 1250 CCGGTATGTC 1259
Db 14 TCGGGGATGC 5

RESULT 2
A2933866 671 bp DNA linear GSS 24-APR-2001
LOCUS Bt_Ba0001M12r B. japonicum BAC library Bradyrhizobium japonicum
DEFINITION genomic, genomic survey sequence.
ACCESSION A2933866
VERSION A2933866.1 GI:13775926
KEYWORDS Bradyrhizobium japonicum
SOURCE Bradyrhizobium japonicum
ORGANISM Bradyrhizobium japonicum
REFERENCE 1 Tomkins,J.P., Wood,T.C., Stacey,M.G., Loh,J.T., Judd,A.,
AUTHORS Goicoechea,J.L., Stacey,G., Sadowsky,M.J. and Ming,R.A.
TITLE A marker-dense, sequence-ready map of the Bradyrhizobium japonicum
genome Res. 11 (8), 1434-1440 (2001)
JOURNAL Genome Res. 11 (8), 1434-1440 (2001)
MEDLINE 21376150
PUBMED 11483585
COMMENT Contact: Wing RA
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4293
Fax: 864 656 4293
Email: twing@clemson.edu
Class: BAC ends
High quality sequence start: 8
High quality sequence stop: 625.
Location/Qualifiers
```

```
source 1..671
/organism="Bradyrhizobium japonicum"
/mol_type="genomic DNA"
/strain="USD110"
/db_xref="taxon:375"
/lab_host="E. coli"
/clone_lib="B. japonicum BAC library"
/notes="Vector: pindigo536; Site_1: HindIII"

Query Match 7.9%; Score 139.8; DB 8; Length 671;
Best Local Similarity 52.7%; Pred.No.1.2e-27;
Matches 323; Conservative 0; Mismatches 289; Indels 1; Gaps 1;

QY 648 AACTTTCGCGGTTTGTGTGAAGCGCGCGGTGTGTGATGATGTCGCGGAGCA 707
Db 616 AACTTTCGCGGTTTGTGTGAAGCGCGCGGTGTGTGATGATGTCGCGGAGCA 557
QY 708 CACCTTCGCGGCGGACCAACAATCGGATCGTGTGTACAAAGCGCGGACGTACTT 767
Db 556 CACATTGGGCGCAACACAAATTCAACCTGCTGTGCGCAAGCGGCGCTCATCTATAT 497
QY 768 TGAACAGGTAAACCAACCGCCCAAGGTTCCGCGAGAGCGCTTCGCACTTGAAGCAAT 827
Db 496 CGAGCGCGCAACCTTCGCGCGCTCTTGTGACGTGCTGCGAAGCTGAAAGCGT 437
QY 828 CTCTCCCACTGCGTACCTCACTGTGCGGAAAGGCTGGAGAAATTAGTGGTCCCTTGA 887
Db 436 GATCGAGACGCTTATTTCAACGTGCGCGGCTTCGACATGCTATACCGGCAATTGGC 377
QY 888 GCGAGCAAGTACCTTCGCGGAAAGCTTCTTTCGCTGCGATGAAAGCTGTTCTTTCGCGC 947
Db 376 CGGTGACGAAGCGTTCGCGCTTCGCTTCAGCGAGGTAAGTTCGCTTTCACGCGG 317
QY 948 GGTGGGTTCGCGCAAGGATCTGGGATCGTTTGAACCGGTCGCGAAGCACTGTG 1007
Db 316 CGGAGCGTCCGCAACAATGTGGACCGCGCTGCACTTATTTCCGTCTTAAACGCTCG 257
QY 1008 TGAAGCATTTGCGATGATGCGGCGGTCTGGGCAATGACGAGACTGCTC-TTCTGCACTT 1066
Db 256 CGGCTTGTCGCGATGATGATGCGGCGGTCTGGGCAATGACGAGACTGCTC-TTCTGCACTT 197
QY 1067 TTAACACCGGACCGCTGTCGATGCTGTGTACATTTGGGCTGCCAGCGCTGCTGAGG 1126
Db 196 GTCCATTTCCTGCGCGAGCTCTCCGTGCAATATCCGCGTTCGATACCCGCTCTGAGC 137
QY 1127 TCAAGCTCGTTCGCGTGTGATGGAATTTGGAAGGGGTTTCATGTTCGCACTGATGA 1186
Db 136 TGAAGCTCGTTCGCGTGTGATGGAATTTGGAAGGGGTTTCATGTTCGCACTGATGA 77
QY 1187 GCGGCTACTGCGCGTGTCTGGAACAATGCCCAAGCTTTCGACGAGAGGCTATTACT 1246
Db 76 TGGGTACTGGAAGGCGCGCGAGCTGACAGGAGGCGTTTCGTCAGAGGCGTTTATC 17
QY 1247 GCTCGGTATGTC 1259
Db 16 TCCTCGTATGTC 4

RESULT 3
A2935488 544 bp DNA linear GSS 24-APR-2001
LOCUS Bt_Ba0003J24r B. japonicum BAC library Bradyrhizobium japonicum
DEFINITION genomic, genomic survey sequence.
ACCESSION A2935488
VERSION A2935488.1 GI:13778288
KEYWORDS Bradyrhizobium japonicum
SOURCE Bradyrhizobium japonicum
ORGANISM Bradyrhizobium japonicum
REFERENCE 1 Tomkins,J.P., Wood,T.C., Stacey,M.G., Loh,J.T., Judd,A.,
AUTHORS
```

RESULT 4					
A2933744/c	A2933744	714 bp	DNA	linear	GSS 24-APR-2007
LOCUS	BU_Ba0001A14r	B. japonicum	BAC library	Bradyrhizobium japonicum	
DEFINITION	genomic, genomic	survey sequence.			
ACCESSION	A2933744				
VERSION	A2933744.1	GI:1375804			
KEYWORDS	GSS.				
SOURCE	Bradyrhizobium japonicum				
ORGANISM	Bradyrhizobium japonicum				

QY	649	ACATTCCCGGTTTTTGGTGAAGAGCGCGGTGCTGGTGAATGGTGGCCGGTGAACAC	708
	652	ACATTCTTGAACATGCTGGACAGCATCTGTCATCTTCGATGAGCTGCCCCGTGAAGCTTC	593
QY	709	ACCTTGGCGCGCAGCCACAACATCGGATGCTGTGTGTACAACGGCGGACGTAACCTT	768
Db	592	ACATTGGGCTTCAACACACATTTTCAACTGTGTATGGGCATCGGGGGTCACTGTATATC	533
QY	769	GACGACGGTAACCAACCGCCCAAGGGTTGCGCAGACGCTTGCGAACTTAGCGAAATC	828
Db	532	GACGGCGGCGAGCGCTGGACCGGTCCTTTGGCAGCTCGCTCGGAACTTACAGCGTG	473
QY	829	TCTTCCACTGCGTAACTTCATCTGTGCGGAAAGGCTGGGAGAAATTAGTGGGTGCCCTTAG	888
Db	472	ATGCCGACGGTCTAATTTTCAACGTGCCCCGGGGTTCCACATGCTGATGCGCTCATTTGCTC	413
QY	889	CGAGACAGTACCTGTGCGCAGACGGTCTTTCGCTCGCATGAAAGCTTTCTTTCGCGGCG	948
Db	412	GGTAGCAAGCGTTGTCTCCCCCGCTTTTCAAGCAGGTGAAGTTTCGCTTTTACGGGGGC	353
QY	949	GCTGGGTTGTGCAAGGGAATCTGGATCTGTTGAGCCGGGTGCTGAACAGACATGTGCT	1006
Db	352	TCAAGCGTGGCGAAGAACTGTGGAAAGCGCCTCTGAATTTGCTGTCATACCGTCCGTC	293
QY	1009	GACGCGATTTGGCATGATGGCGGGGTCTGGGATGACGAGACTGCTCTTCTTGACATTTT	1066
Db	292	CGCCCGCTGGCATGTATGCCCCCTGGGGCTTCGACGGAACCTGCGCGCTGGCGACCGCC	233
QY	1069	ACCACCGGACCGGATGCTGATGGCTGTATACATTTGGGCTGCGACGCGCTGGCGAGGTC	1128
Db	232	TGCGATTTTCTTCCGCCGACGCGCTCCGTCAATATCGCGCGTGGCAATCCCGTATACGAGCTG	173
QY	1129	AAGCTGTTTCCGTCGATGGTAGTGAATTTGAAAGGCGTTTCCATGTCGCGACGTATGAGC	1188
Db	172	AAGCTGTCACCTCCGTCGCCAAAGCTGATGTGCGCGTGGCGGCCGCCACCGCTCTCGCCA	113
QY	1189	GGCT 1192	

Db	112	GTT	109
RESULT 5	CNS0175Y	1101 bp	DNA
CNS0175Y/c			linear GSS 26-JUL-1999
LOCUS	Drosophila melanogaster genome survey sequence SP6 end of BAC		
DEFINITION	BACN37108 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.		
ACCESSION	AL108460		
VERSION	AL108460.1		
KEYWORDS	GI:5628764		
SOURCE	GSS.		
ORGANISM	Drosophila melanogaster (fruit fly)		
REFERENCE	Drosophila melanogaster		
AUTHORS	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		
TITLE	1 (bases 1 to 1101)		
JOURNAL	Genoscope.		
COMMENT	Direct Submission Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : sequefgenoscope.cns.fr - Web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (BDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.		
FEATURES	Location/Qualifiers		
source	1..1101		
	/organism="Drosophila melanogaster"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:7227"		
	/clone="BACN37108"		
	/clone_11b="DrosBAC"		
	/plasmid="pBelobAC11"		
	/note="end : SP6"		
ORIGIN			
Query Match	2.7%: Score 47.4: DB 9: Length 1101:		
Best Local Similarity	19.1%: Pred. No.0.11,		
Matches	72: Conservative 142: Mismatches 162: Indels 0: Gaps 0:		
Qy	793 GGGTTCGCCGAGACGCTTCGCACTTGAGCGAATCTCTCCACATCGCTCACTGTG	852	
Db	986 BSSBSBSBSSTSSBRTSKBTSBBSBSTSSABBSBSBSBSSTSTTSBBSBTS	927	
Qy	853 CCGAAGCGCTGGAGAAATTAGTGGTCCCTTGACGACAGACAGTACCTCGCGCAACG	912	
Db	926 SSSGSSSSBSBSBTSBBSBRTTTTBTKSTSTSSSTSBTTTSBSSBSBSBSB	867	
Qy	913 TTCTTCGCTCCATGAAAGCTGTTCTTCTTCGCGCGCGCTGGATTGCCAAGGATCTGG	972	
Db	866 STTTSBSBTSBSYSSBSBTSKSTBTSTGTSBTSBKSBSSTSTBTSBSBSCGSSC	807	
Qy	973 GATCGTTTGACCGGGCTGCTGAACAGACACTGTGTGAGCGGATTGCATGATGGCGGTT	1032	
Db	806 GSTGSSCSBSBCKSTSSGTSSTGTTCCBSCSTGACSCCCCTCCCTCTGATGTCGCS	747	
Qy	1033 CTGGGCAATGACGAGACGCTGCTCTTCCTCGACCTTTTACCGGACCGCTGTCATGTGCT	1092	
Db	746 CSGSSBSBKCYSTSSBSASTGSSCSSTSSSGGCTGTCGCGATTKGTGTCCTCGBY	687	
Qy	1093 GGTTCATTGGAGCTGCGACCGCTGCTGCGACGATCAAGTCTGTTCCGATGGAGAA	1152	
Db	686 CCYBCCTMTMCCGSSCCSGSYCGSSSKCTSMGTTGCMYCCGCTSCCGTGAGGAMA	627	
Qy	1153 TTGGAAGGCGCTTCC	1168	
Db	626 CGTGCCTGSGCGNKNVM	611	

[illegible]

REFERENCE	AUTHORS	TITLE
1	(bases 1 to 660)	Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchima, Tylenchoidea; Heterodermidae; Heterodermiae; Heterodera.
2	Gao, B., Allen, R., Meier, T., Davis, E. L., Baum, T. J. and Huesey, R. S.	Identification of putative parasitism genes expressed in the esophageal gland cells of the soybean cyst nematode, <i>Heterodera glycines</i>
3	Mo, Plant Microbe Interact.	14 (10), 1247-1254 (2001)
4	Contact: Tom Meier Department of Plant Pathology, Baum Lab Iowa State University 351 Bessey Hall, Ames, IA 50011, USA Tel: 515-294-8854 Fax: 515-294-9420 Email: tmeier@iastate.edu	Heterodera glycines Gland Cell LD PCR cDNA Library, single pass sequence.
5	Location/Qualifiers	1..660
6	/organism="Heterodera glycines"	
7	/mol_type="mRNA"	
8	/db_xref="taxon:51029"	
9	/cell_type="gland"	
10	/dev_stage="mixed parasitic juvenile"	
11	/clone_lib="Gland Cell LD PCR cDNA Library"	
12	/note="Organ: gland cell; Vector: pGEM-T Easy"	
13	Query Match	2.4%; Score 42.8; DB 7; Length 660;
14	Best Local Similarity	57.5%; Pred. No. 1.8;
15	Matches	77; Conservative 0; Mismatches 57; Indels 0; Gaps 0;
16	QY	1136 TTCCGCTCGATGGAAATTGGAAGGCGCTTTCATGTCGCCACGTCATGAGCGGCTACT 1195
17	Db	4 TTACAGATGATCGGAAGGCGCAAGTGTGTGGAAGGAGCCGTCGTGACCAAGGCTATT 63
18	QY	1196 GCGGTGCTCTCGAACAAATGCCCCAAGCGTTGACGAGAGAGCTATTACTGCTCCGGT 1255
19	Db	64 ACCGGAGCGCGGAAAGACCGCGCATTTGTCATGAGAGACGCGCTTTTGGACACCGGAG 123
20	QY	1256 ATGCCATCAATTG 1269
21	Db	124 ACATTGGCCCAATTG 137
22	RESULT 8	
23	LOCUS	BX321259 585 bp mRNA linear EST 20-MAY-2004
24	DEFINITION	BX321259 tcay Oncorhynchus mykiss cDNA clone tcay0024b.n.02 5'prim, mRNA sequence.
25	ACCESSION	BX321259
26	VERSION	BX321259.2 GI:42626956
27	KEYWORDS	EST.
28	SOURCE	Oncorhynchus mykiss (rainbow trout)
29	ORGANISM	Oncorhynchus mykiss Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
30	REFERENCE	1 (bases 1 to 585) Govoroun, M., Guiguen, Y. and Le Gac, F.
31	AUTHORS	Construction and primary characterization of normalized cDNA libraries in rainbow trout, <i>Oncorhynchus mykiss</i>
32	TITLE	Unpublished (2003)
33	JOURNAL	On Apr 7, 2003 this sequence version replaced gi:29601904.
34	COMMENT	Contact: Guiguen Y INRA - SCRIBS Campus de Beaulieu, RENNES cedex, 35042, France Tel: 02.23.48.50.09 Fax: 02.23.48.50.20 Email: Yann.Guiguen@beaulieu.rennes.inra.fr Sequence cleaned of vector, adaptor and repetitions. Contact us at eigenasupport@jouy.inra.fr to obtain the chromatogram of this sequence.

FEATURES	Source	Location/Qualifiers
ORIGIN		1..585 /organism="Oncorhynchus mykiss" /mol_type="mRNA" /db_xref="taxon:8022" /clone="tcay0024b.n.02" /tissue_type="adipose tissue, blood, brain, differentiating gonads, gills, interrenal, intestine, kidney, liver, muscle, ovary, pituitary, testis" /dev_stage="from embryos to adults" /lab_host="DH108" /clone_lib="tcay" /note="Vector: pT7T3D-pac; Rainbow trout multi-tissues - normalized + 1 subtraction (tcay) ; Clone distribution : AGENAE Resource centre, Francois PIUMI, Francois.Piumi@jouy.inra.fr, INRA, CEA Radiobiologie et Etude du genome (LREG), Domaine de Vilvert, 78352, Jouy-en-Josas cedex, France, +33 (0) 1.34.65.28.02, +33 (0) 1.34.65.22.73"
Query Match	2.4% Score 42.6; DB 5; Length 585;	
Best Local Similarity	51.9% Pred. No. 2;	
Matches	96; Conservative 0; Mismatches 89; Indels 0; Gaps 0;	
OY	1085 CGATGCGTGTATCATTTGGGCTGCAGGCGCTGCGAGTCAAGCTCGTCCGTCG	11444
DB	165 CTATACCTGTGAACATTGTGAAACGTGTGATGTGAAGAATGATGAATCTTGTCTCA	224
OY	1145 ATGGAAATTGGAAGGCGTTCCATGTCCGCACGTATAGAGGCGCTACTGGCTGCTC	12044
DB	225 ATGTGTAAGGGAGAGGTCTGTATCAAAGGTAAACGTGTTATCGAATCTGAAAGAC	284
OY	1205 CTGAACAATAATGCCAGCGTTGACAGCGAAGGCTATTACTGCTCGGCGTATGCATCA	12644
DB	285 CTGAGAAAGACGACAGAGGCGCTGACAAAGATGCTGCTACACACGGAGACATTGGCA	344
OY	1265 AATTG 1269	
DB	345 AATGG 349	
RESULT 9		
EX882027		
LOCUS	795 bp mRNA linear EST 20-JUL-2004	
DEFINITION	EX882027 tcdh Oncorhynchus mykiss cDNA clone tcdh0055c.k.18 5prim,	
ACCESSION	EX882027	
VERSION	EX882027.2 GI:42814986	
KEYWORDS	EST.	
SOURCE	Oncorhynchus mykiss (rainbow trout)	
ORGANISM	Oncorhynchus mykiss	
REFERENCE	Euhayot, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus. 1 (bases 1 to 795)	
AUTHORS	Govoroun, M., Guiguen, Y. and Le Gac, F.	
TITLE	Construction and primary characterization of normalized cDNA libraries in rainbow trout, Oncorhynchus mykiss	
JOURNAL	Unpublished (2003)	
COMMENT	On Dec 18, 2003 this sequence version replaced gi:40126280. Contact: Guiguen Y INRA - SCRIBE Campus de Beaulieu, RENNES cedex, 35042, France Tel: 02.23.48.50.09 Fax: 02.23.48.50.20 Email: Yann.Guiguen@beaulieu.rennes.inra.fr Sequence cleaned of vector, adaptor and repetitions. Contact us at sigenasupport@jouy.inra.fr to obtain the chromatogram of this sequence.	
row: k	column: 18	

```

FEATURES
source
Seq primer: M13r.
location/Qualifiers
1..795
/organism="Oncorhynchus mykiss"
/mol_type="mRNA"
/db_xref="taxon:8022"
/clone="tcdbK0053c.k.18"
/tissue_type="multi-tissues"
/dev_stage="from embryos to adults"
/lab_host="DH10B"
/clone_lib="tcdbk"
/note="vector: pRTT3D-pac; AGENAE Rainbow trout
multi-tissues - normalized + 2 subtractions; Clone
distribution: AGENAE Resource centre, Francois Piumi,
Francois.Piumi@jouy.inra.fr, INRA, CEA Radiobiologie et
Etude du genome (IREG), Domaine de Vilvert, 78352,
Jouy-en-Josas cedex, FRANCE, +33 (0) 1.34.65.28.02, +33
(0) 1.34.65.22.73"

```

Query Match	2.4%	Score 42.6	DB 5	Length 795
Best Local Similarity	51.9%	Pred. No. 2.1		
Matches	96	Conservative	0	Mismatches 89
				Indels 0
				Gaps 0
QY	1085	CGATGGCTGTATTACATTGGGCTGCGCAGCGCTGCTGGAGGTCAAGCTGTTCCGTCG	1144	
Db	72	CTATACCTTTGTAACATTGTGAAACTGTTGATTGAAGAGATGAATCACTTGGCTCA	131	
QY	1145	ATGGGAATTGGAAAGGCGCTTTCATGCTCCGCAAGTATGAGCGGTACTAGCGCTGCT	120	
Db	132	ATGGTGAAGGGGAGGTCTGTATCAAAAGGTAAAAAGCTTCAACGAAATCTGAAAGACC	191	
QY	1305	CTGAACAATAATGCCCAAGCGTTTCGACGAGAAAGCTATTACTGCTCCGGTATGTCATCA	126	
Db	192	CTGAGAAACACGACAGAGCGCTTCGACAAAGATGCTGCTACACACCGGAGACATTGGCA	251	
QY	1265	AATTG 1269		
Db	252	AATGG 256		

RESULT 10	
BX911066	
LOCUS	BX911066 815 bp mRNA linear EST 27-Jul-2004
DEFINITION	BX911066 tcbk Oncorhynchus mykiss cDNA clone tcbk0068c.d.13 5'prim, mRNA sequence.
ACCESSION	BX911066
VERSION	BX911066.2 GI:43416323
KEYWORDS	EST.
SOURCE	Oncorhynchus mykiss (rainbow trout)
ORGANISM	Oncorhynchus mykiss
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus. 1 (bases 1 to 815) Govoroun, M., Guiguen, Y. and Le Gac, F. Construction and primary characterization of normalized cDNA libraries in rainbow trout, <i>Oncorhynchus mykiss</i> unpublished (2003)
JOURNAL	
COMMENT	On Jan 22, 2004 this sequence version replaced gi:41127874.

Campus de beaulieu, RENNES cedex, 35042, France
Tel : 02.23.48.50.09
Fax: 02.23.48.50.20
Email: Yann.Guiguen@beaulieu.rennes.inra.fr
Sequence cleaned of vector, adaptor and repetitions. Contact us
at sienasupport@joely.inra.fr to obtain the chromatogram of this
sequence.
Plate: 0068 row: d column: 13
Seq primer: M13R
Location/Qualifiers
1. 815

```

/organism="Oncorhynchus mykiss"
/mol_type="mRNA"
/db_xref="taxon:8022"
/clone="tcdb0068c.d.13"
/tissue_type="multi-tissues"
/dev_stage="from embryos to adults"
/lab_host="DH10B"
/clone_lib="tcdbk"
/note="vector: pRT73D-pac; AGENAE Rainbow trout
multi-tissues - normalized + 2 subclonings; Clone
distribution : AGENAE Resource centre, Francois PIUMI,
Francois.PIUMI@jouy.inra.fr, INRA, CEA Radiobiologie et
Etude du genome (IREC), Domaine de Vilvert, 78352,
Jouy-en-Josas cedex, FRANCE, +33 (0) 1.34.65.28.02, +33
(0) 1.34.65.22.73"

```

Oy	CGATGGCGTGTTCATTATGAGCTGCACAGGCCCTGGCTGGAGATCAAGCTCCTCCGGTGC	1144
Db	CTAATACCTTGTAACATTGTGAAACTGGTTGATGTTGAAGAATGAACTACTTGGCTTCCA	418
Oy	ATGGGAATTTGGAAAGGCGGTTTTCCATGATGCCACGTCATAGCGCGCTACTGGCGTCTC	1204
Db	ATGTGGAAGGGGAAGSTCTGTATCAAAGTAAAAAGTGTTCATGGATATATCGAAAAGACC	478
Oy	CTGAACAATAATGCCCAAGCGTTGCACGACGAGAAGGCTATTACTGCTCCGGTAGATGCATCA	1264
Db	CTGGAAGACACGACAGAGGCCCTCGAACAAAGATGTGCTGTACACACCGGAGACATTGCCA	538
Oy	AATTGG	1269
Db	AATGG	543

RESULT 11	
CNS004NB	
LOCUS	CNS004NB 839 bp DNA linear GSS 03-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence TIR3 end of BAC # BACR10616 of RPc1-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION	AL054280
VERSION	AL054280.1 GI:4931788
KEYWORDS	GSS.
SOURCE	Drosophila melanogaster (fruit fly)
ORGANISM	Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephyridroidea; Drosophilidae; Drosophila. 1 (bases 1 to 839) Genoscope. Direct Submission Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr - Web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
COMMENT	

TITLE	COMMENT
<p>Direct Submission Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr)</p>	<p>- Web : www.genoscope.cns.fr - Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.flytiffy.org/The_BDGP_Drosophila_melanogaster_BAC_library was prepared by Kazuhiro Osoegawa and Aaron Mammoeser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPc1-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.</p>

FEATURES	source	location/Qualifiers
1. 839		location/Qualifiers
/organism="Drosophila melanogaster"		
/mol_type="genomic DNA"		
/db_xref="taxon:7227"		
/clone="BACRI0E16"		
/clone_11b="RPC1-98"		
/note="end : TET3"		
Query Match	2.3%; Score 41.4; DB 9; Length 839;	
Best Local Similarity	18.4%; Pred. No. 4.7;	
Matches	69; Conservative 129; Mismatches 176; Indels 0; Gaps 0;	
QY	746 ACAACGGCGGCACTACTTGAACGACGCGTAACCAACCGCCCAAGGTTCCCGCAGA	805
Db	430 ANNACCCACMCCNNATATNTTATTAAGAAATSMSSSACSAASGCSASCVAA	489
QY	806 CGCTTCGCACTTGAGCGAAATCTCTCCACTCGCTACTCTACTGTGCGAAAGGCTGG	865
Db	490 CSSACAGSAGSAGSSSSASGCGAAGSSSAGSGAGSRRGSGGAGSGGSGGGSRRSG	549
QY	866 AGGATTAAGTGGGTGCGCCCTTGAGCGAGACAGTACCCTGCGGACGCTTCTTCGCTCGA	925
Db	550 CGGAVGASAGSSSASGCSASSASSSSASGCMSCSSSSSSSSASGSSSSSGVAATCS	609
QY	926 TGAAGCTGTTCTTCTTCGCGCGCGCTGGGTGTGTCGCAAGGATTCGGATCGTTTGACC	985
Db	610 GSGGAGVGSAAVAASVARSAGRGCGAAGSRAVASAAASASASVAVGAACSGS	669
QY	986 GCGTCGCTGAACAGCATCTGTGTAGCGCCATTCGCATGATGCGGCTCTGGCATGACG	1045
Db	670 GASGSSSSGAARARCAVCASSASSGSCGSSSVCASVSSGASGASGCCGCGSSSG	729
QY	1046 AGACTGCTCTTCCTGCACTTTTACCAACGACCGCTGCTGATGCGCTTACATTGGGC	1105
Db	730 CSGCSCSSSSSCSSSSCGSSCGSSCGSSCGSSCGSSCGSSCGSSCGSSCGSSCGSS	789
QY	1106 TGCCAGCGCGCTGCG 1119	
Db	790 SSSCSVSCSMSC 803	
RESULT 12		
LOCUS	BM442982	
DEFINITION	EBR002_SB002_B12_R rooc. 3 week, hydroponic grown, low nitrogen, cv	
VERSION	EBR002_SB002_B12 5', mRNA sequence.	
KEYWORDS	EBR002_SB002_B12 5', mRNA sequence.	
ACCESSION	BM442982	
VERSION	BM442982.2 GI:21947637	
KEYWORDS	EST.	
SOURCE	Hordeum vulgare subsp. vulgare	
ORGANISM	Hordeum vulgare subsp. vulgare	
REFERENCE	Bukartova, V. Iridiplantae; Striptophyta; Embryophyta; Tracheophyta;	
AUTHORS	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;	
TITLE	Pooidae; Trilicaceae; Hordeum.	
JOURNAL	1 (bases 1 to 240)	
COMMENT	Healey, P., Liu, H., Caldwell, D., McCallum, N., Mudie, S., Cardie, L.,	
	Ramsey, L., Mchray, G., Marshall, D.F.M. and Waugh, R.	
	Development of barley transcriptome Resources	
	Unpublished (2001)	
	On Feb 1, 2002 this sequence version replaced gi:18473745.	
	Contact: Waugh R, Marshall DF	
	Genome Dynamics/Computational Biology	
	Scottish Crop Research Institute	
	Invergowrie, Dundee, DD2 5DA, Scotland, UK	
	Tel: 00 44 1382 562731	
	Fax: 00 44 1382 562426	
	Email: est@scri.sari.ac.uk	
	All sequence has a phred quality score of 20 or over	
	Seq primer: M13 reverse	
	Location/Qualifiers	

1. .240
 /organism="Hordeum vulgare subsp. vulgare"
 /mol_type="mRNA"
 /cultivar="Optic"
 /sub_species="vulgare"
 /db_xref="taxon:112509"
 /clone="EBR002_SQ002_B12"
 /issue_type="root"
 /dev_stage="3 week"
 /lab_host="DH10B"
 /clone_1fb="root, 3 week, hydropenic grown, low nitrogen, cv Optic, EBR002"
 /note="Vector: pSPORT1; Site 1: Sal I; Site 2: Not I; Non-normalised library, directionally cloned into pSPORT1. Derived from roots of 3 week old Nitrogen stressed barley plants. Developed as part of the barley transcriptome resources of BBSRC/SEERAD funded cereal IGF (Investigating Gene Function) project."

Query Match 2.3%; Score 41.2; DB 4; Length 240;
 Best Local Similarity 56.7%; Pred. No. 3.9;
 Matches 76; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

OY 551 TTGCCAGTCTTGTTCACCTTGTGGCTTCACCAATCGCTAAGGCGGTGCGCACTACTC 610
 39 TGGCCCTGTCTGCACACCTCCGCGACAGACCCGCCCMAGGTGTGCTGCGC 98

OY 611 AGCGAATGCTCTGCGCCCAATCAGCAGATGCTTCGCAAACTTCCCGGTTTGTGTAG 670
 99 AGCGGAACCTCGCGCTCGCTCGTCCAGAACATCCGCGGTGACCGGTTGTGAGACGG 158

OY 671 AGCCGCGGTGCTG 684
 DB 159 ACGCAGCGGTGCTG 172

RESULT 13
 LOCUS EBR002_SQ002_D13_R root, 3 week, hydropenic grown, low nitrogen, cv
 EBR002_SQ002_D13_5', mRNA sequence.
 ACCESSION BM443022
 VERSION BM443022.2 GI:21947674
 KEYWORDS EST.
 SOURCE Hordeum vulgare subsp. vulgare
 ORGANISM Hordeum vulgare subsp. vulgare
 Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 1 (bases 1 to 363)
 Hedley, P., Liu, H., Caldwell, D., McCallum, N., Mudie, S., Cardle, L.,
 Ramsay, L., Machray, G., Marshall, D.F.M. and Maugh, R.
 Development of Bailey Transcriptome Resources
 Unpublished (2001)
 On Feb 1, 2002 this sequence version replaced gi:18473785.
 Contact: Maugh R, Marshall DP
 Genome Dynamics/Computational Biology
 Scottish Crop Research Institute
 Invergowrie, Dundee, DD2 5DA, Scotland, UK
 Tel: 00 44 1382 562731
 Fax: 00 44 1382 562426
 Email: est@scri.sari.ac.uk
 All sequence has a Phred quality score of 20 or over
 Seq primer: M13 reverse.

FEATURES
 Location/Qualifiers
 1..363
 /organism="Hordeum vulgare subsp. vulgare"
 /mol_type="mRNA"
 /cultivar="Optic"
 /sub_species="vulgare"
 /db_xref="taxon:112509"

/clone="EBR002_S0002_D13"
/tissue_type="root"
/dev_stage="3 week"
/lab_host="DH103"
/clone_lib="root", 3 week, hydroponic grown, low nitrogen,
cv Optic, EBR002"
/note=Vector: pSPORT1; Site 1: Sal I; Site 2: Not I;
Non-normalised library, directionally cloned into pSPORT1.
Derived from roots of 3 week old Nitrogen stressed barley
plants. Developed as part of the barley transcriptome
resources of BBSRC/SEERAD funded cereal IGF (Investigating
Gene Function) Project."

ORIGIN

Query Match 2.3%; Score 41.2; DB 4; Length 363;
Best Local Similarity 56.8%; Pred. No. 4.3;
Matches 76; Conservative 0; Mismatches 58; Indels 0; Gaps 0;
QY 551 TTGCCAAGTTCTTGTTCATTCTTGGCTCTTACCAAACTGCTTAAAGCGGTGCCGACTACTC 610
DB 39 TGCCCTGTTCTGCACACCTCCGCGCAGAGCCGCAAGGAGTGTCCGCTGACGC 98
QY 611 AGGAATGCTCTGCGGCAATCAGAGATGCTTCTGCAACTTCCCGTTTGTGTAAG 670
DB 99 AGGCAACCTCGCGCCTCCGTCACAGAACATCCGCGCTGTACCGTTGTGAGACGG 158
QY 671 AGCGCGCGTGTG 684
DB 159 ACGCGACGTGTG 172

RESULT 14
CR676208 856 bp mRNA linear HTC 12-AUG-2004
LOCUS Tetraodon nigroviridis full-length cDNA.
DEFINITION CR676208.1 GI:51173964
VERSION CR676208.1
KEYWORDS HTC; cDNA; full-length; Tetraodon nigroviridis.
SOURCE Tetraodon
ORGANISM Tetraodon
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
Tetraodontidae; Tetraodontidae.
REFERENCE 1 (bases 1 to 856)
Genoscope.
AUTHORS Direct Submission
TITLE Submitted (10-AUG-2004) Genoscope - Centre National de Sequencage -
JOURNAL : 2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
(E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT The sequences are based on single pass reads.
More information available at
http://www.genoscope.cns.fr/tetraodon.
FEATURES
source
1..856
/organism="Tetraodon"
/mol_type="mRNA"
/db_xref="taxon:47144"
/tissue_type="Muscle"

ORIGIN

Query Match 2.3%; Score 40.8; DB 3; Length 856;
Best Local Similarity 56.8%; Pred. No. 7;
Matches 75; Conservative 0; Mismatches 57; Indels 0; Gaps 0;
QY 448 GCGGAATGGCCGGCGCGCAGCGTGAATTTGACAGCCTGCTGAGCAGCCTGTGGG 507
DB 657 GTCAAGATGCTCAAGAGCCCAAGTTGAGTTGGCCAACTGATGAGCTCCACGGTGA 716
QY 508 ATTGAGCAGATATATGCTTTGGCGCAACTGCCCCGATACGATTCGCAAGTTCTTGTTC 567
DB 717 GGTGGGCGAGAACTGCTGCTAAGGCACTGTGACGATTACTGAGCCCAAGTGGACGG 776

QY 568 ACTTCTGGCTCT 579
DB 777 GCTGATGGCTAT 788

RESULT 15
CR662673 873 bp mRNA linear HTC 11-AUG-2004
LOCUS Tetraodon nigroviridis full-length cDNA.
DEFINITION CR662673
VERSION CR662673.1 GI:51159118
KEYWORDS HTC; cDNA; full-length; Tetraodon nigroviridis.
SOURCE Tetraodon
ORGANISM Tetraodon
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
Tetraodontidae; Tetraodontidae.
REFERENCE 1 (bases 1 to 873)
Genoscope.
AUTHORS Direct Submission
TITLE Submitted (10-AUG-2004) Genoscope - Centre National de Sequencage -
JOURNAL : 2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
(E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT The sequences are based on single pass reads.
More information available at
http://www.genoscope.cns.fr/tetraodon.
FEATURES
source
1..873
/organism="Tetraodon"
/mol_type="mRNA"
/db_xref="taxon:47144"
/tissue_type="Muscle"

ORIGIN

Query Match 2.3%; Score 40.8; DB 3; Length 873;
Best Local Similarity 56.8%; Pred. No. 7;
Matches 75; Conservative 0; Mismatches 57; Indels 0; Gaps 0;
QY 448 GCGGAATGGCCGGCGCGCAGCGTGAATTTGACAGCCTGCTGAGCAGCCTGTGGG 507
DB 666 GTCAAGATGCTCAAGAGCCCAAGTTGAGTTGGCCAACTGATGAGCTCCACGGTGA 725
QY 508 ATTGAGCAGATATATGCTTTGGCGCAACTGCCCCGATACGATTCGCAAGTTCTTGTTC 567
DB 726 GGTGGGCGAGCACTGCTGCTAAGGCACTGTGACGATTACTGAGCCCAAGTGGACGG 785
QY 568 ACTTCTGGCTCT 579
DB 786 GCTGATGGCTAT 797

Search completed: February 7, 2005, 21:35:19
Job time : 5109 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2005, 20:09:50 ; Search time 979 Seconds

(Without alignments)
10388.378 Million cell updates/sec

Title: US-09-750-986D-29

Perfect score: 1770
Sequence: 1 atgcgcctctcgcagcgcgc.....gtgcagagccacactgtga 1770

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4300275 seqs, 2872944193 residues

Total number of hits satisfying chosen parameters: 8600550

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 100%
Listing first 45 summaries

Database :

Published Applications NA: *
1: /cgn2_6/prodata/2/pubpna/US07_PUBCOMB.seq: *
2: /cgn2_6/prodata/2/pubpna/US09C_PUBCOMB.seq: *
3: /cgn2_6/prodata/2/pubpna/US06_PUBCOMB.seq: *
4: /cgn2_6/prodata/2/pubpna/US07_PUBCOMB.seq: *
5: /cgn2_6/prodata/2/pubpna/US08_PUBCOMB.seq: *
6: /cgn2_6/prodata/2/pubpna/US08_PUBCOMB.seq: *
7: /cgn2_6/prodata/2/pubpna/US08_PUBCOMB.seq: *
8: /cgn2_6/prodata/2/pubpna/US08_PUBCOMB.seq: *
9: /cgn2_6/prodata/2/pubpna/US09A_PUBCOMB.seq: *
10: /cgn2_6/prodata/2/pubpna/US09B_PUBCOMB.seq: *
11: /cgn2_6/prodata/2/pubpna/US09C_PUBCOMB.seq: *
12: /cgn2_6/prodata/2/pubpna/US09C_PUBCOMB.seq: *
13: /cgn2_6/prodata/2/pubpna/US10A_PUBCOMB.seq: *
14: /cgn2_6/prodata/2/pubpna/US10B_PUBCOMB.seq: *
15: /cgn2_6/prodata/2/pubpna/US10C_PUBCOMB.seq: *
16: /cgn2_6/prodata/2/pubpna/US10D_PUBCOMB.seq: *
17: /cgn2_6/prodata/2/pubpna/US10E_PUBCOMB.seq: *
18: /cgn2_6/prodata/2/pubpna/US10F_PUBCOMB.seq: *
19: /cgn2_6/prodata/2/pubpna/US11_PUBCOMB.seq: *
20: /cgn2_6/prodata/2/pubpna/US60_PUBCOMB.seq: *
21: /cgn2_6/prodata/2/pubpna/US60_PUBCOMB.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1770	100.0	1770	8	US-08-976-063C-29
2	1770	100.0	1770	8	US-09-750-986D-29
3	1770	100.0	32679	10	US-08-976-063C-1
4	1770	100.0	32679	10	US-09-750-986D-1
5	527.8	29.8	1725	15	US-10-369-493-44177
6	398.4	22.5	1728	15	US-10-369-493-35384
7	398.4	22.5	1728	15	US-10-369-493-38427
8	398.4	22.5	1728	15	US-10-369-493-38803
9	372	21.0	1660	15	US-10-369-493-38007
10	225.6	12.7	1866	15	US-10-369-493-40734
11	57.4	3.2	1879	18	US-10-716-803-5
12	46.8	2.6	536	16	US-10-338-110-119

13	46.8	2.6	1566	15	US-10-369-493-42307	Sequence 42307, A
14	46.4	2.6	1644	16	US-10-282-122A-30430	Sequence 30430, A
15	45.4	2.6	6522	16	US-10-282-122A-31533	Sequence 31533, A
16	45.2	2.6	1977	15	US-10-369-493-37463	Sequence 37463, A
17	44.4	2.5	1632	16	US-10-282-122A-18967	Sequence 18967, A
18	43.4	2.5	342	16	US-10-282-122A-10967	Sequence 10967, A
19	43.4	2.5	410	16	US-10-282-122A-3010	Sequence 3010, A
20	43.4	2.5	1701	15	US-10-369-493-24535	Sequence 24535, A
21	43.4	2.5	1701	16	US-10-282-122A-6822	Sequence 6822, A
22	42.4	2.4	1677	15	US-10-369-493-41695	Sequence 41695, A
23	42	2.4	1701	15	US-10-369-493-41079	Sequence 41079, A
24	41.6	2.4	825	17	US-10-437-963-73949	Sequence 73949, A
25	41.6	2.4	882	14	US-10-184-644-574	Sequence 574, A
26	41.6	2.4	882	14	US-10-184-644-574	Sequence 574, A
27	41.4	2.3	1680	15	US-10-369-493-33018	Sequence 33018, A
28	41.4	2.3	1680	15	US-10-369-493-33087	Sequence 33087, A
29	41.4	2.3	1803	17	US-10-437-963-43352	Sequence 43352, A
30	40.6	2.3	1500	15	US-10-369-493-43198	Sequence 43198, A
31	40.6	2.3	13029	9	US-09-815-242-4052	Sequence 4052, A
32	40.6	2.3	13029	16	US-10-282-122A-7240	Sequence 7240, A
33	40.4	2.3	1049	14	US-10-123-155-358	Sequence 358, A
34	40.4	2.3	1049	15	US-10-146-731-358	Sequence 358, A
35	40.4	2.3	1049	15	US-10-140-472-358	Sequence 358, A
36	40.4	2.3	1049	15	US-10-141-761-358	Sequence 358, A
37	40.4	2.3	1049	15	US-10-142-885-358	Sequence 358, A
38	40.4	2.3	1049	15	US-10-158-790-358	Sequence 358, A
39	40.4	2.3	1049	15	US-10-137-821-358	Sequence 358, A
40	40.4	2.3	1049	15	US-10-140-923-358	Sequence 358, A
41	40.4	2.3	1049	15	US-10-141-756-358	Sequence 358, A
42	40.4	2.3	1049	15	US-10-141-759-358	Sequence 358, A
43	40.4	2.3	1049	15	US-10-140-805-358	Sequence 358, A
44	40.4	2.3	1049	15	US-10-140-864-358	Sequence 358, A
45	40.4	2.3	1049	16	US-10-142-426-358	Sequence 358, A

ALIGNMENTS

RESULT 1
US-08-976-063C-29
; Sequence 29, Application US/08976063C
; Publication No. US20020182697A1
GENERAL INFORMATION:
APPLICANT: Alexander Steinhuechel; Horst Prietert; Jürgen Rabenhorst
TITLE OF INVENTION: SYNTHETIC ENZYMES FOR THE PRODUCTION OF
TITLE OF INVENTION: CONFERTAL ALCOHOL, CONFERTALDEHYDE, FERULIC ACID, VANILLIN A
TITLE OF INVENTION: ACID AND THEIR USE
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESSES:
ADDRESSEE: SPRING KRAMER SCHAEFER & BRISCOE
STREET: 660 White Plains Road
CITY: Tarrytown
STATE: New York
COUNTRY: U.S.A.
ZIP: 10591-5144
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB storage
COMPUTER: HP VECTRA
OPERATING SYSTEM: DOS
SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/976, 063C
FILING DATE: 21-NOV-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 196 49 655.1 (Germany)
FILING DATE: 29-NOV-1996
ATTORNEY/AGENT INFORMATION:
NAME: Kurt G. Briscoe
REGISTRATION NUMBER: 33,141
REFERENCE/DOCKET NUMBER: Bayer 9998-CAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (914) 332-1700

TELEFAX: (914) 332-1844
TELEX:
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 1770 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHEetical: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1767
OTHER INFORMATION: /product=
OTHER INFORMATION: "Ferulaseure-CoA-Synthetase"
OTHER INFORMATION: /gene="fcs"
US-08-976-063C-29

Query Match 100.0%; Score 1770; DB 8; Length 1770;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1770; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCGTCTCTCGAGGCGCTTCTTCCCTTCCCGGCTGAATTCTTGAGCGCTTCGAGCAT 60
DB 1 ATGCGTCTCTCGAGGCGCTTCTTCCCTTCCCGGCTGAATTCTTGAGCGCTTCGAGCAT 60
QY 61 TGGGCTAAGACCGCTCCAGAACTGCGGCTGCTGCGGCGGCGAAATGGGGAATGG 120
DB 61 TGGGCTAAGACCGCTCCAGAACTGCGGCTGCTGCGGCGGCGAAATGGGGAATGG 120
QY 121 CGTGTATCACTAAGCGGAAATGTTTCACAACTGCGGCGCAATGCGACAGAGCTTCTT 180
DB 121 CGTGTATCACTAAGCGGAAATGTTTCACAACTGCGGCGCAATGCGACAGAGCTTCTT 180
QY 181 CTTTACGAGATATGCGAGAGCGCTGCTGCTTATCGTCTCTGGAATGACCTGGAACAT 240
DB 181 CTTTACGAGATATGCGAGAGCGCTGCTGCTTATCGTCTCTGGAATGACCTGGAACAT 240
QY 241 CTTTACGAGATATGCGAGAGCGCTGCTGCTTATCGTCTCTGGAATGACCTGGAACAT 240
DB 241 CTTTACGAGATATGCGAGAGCGCTGCTGCTTATCGTCTCTGGAATGACCTGGAACAT 240
QY 301 TATTTCACTGCTGCGCAAGATTGGCGAAGCTGCTGCAATCGTAGGTCTTTGCAACCG 360
DB 301 TATTTCACTGCTGCGCAAGATTGGCGAAGCTGCTGCAATCGTAGGTCTTTGCAACCG 360
QY 361 GGAAGTCTCTTGTGCGGATGAGCACTTTTCCAGGCGCAATGGAACCAATTTGCGC 420
DB 361 GGAAGTCTCTTGTGCGGATGAGCACTTTTCCAGGCGCAATGGAACCAATTTGCGC 420
QY 421 GAGGAGCTGCGCGCAATCTTCACTGAGGCGAATTTGCGGCGCGCAACGTTGCGT 480
DB 421 GAGGAGCTGCGCGCAATCTTCACTGAGGCGAATTTGCGGCGCGCAACGTTGCGT 480
QY 481 GAGGAGCTGCGCGCAATCTTCACTGAGGCGAATTTGCGGCGCGCAACGTTGCGT 540
DB 481 GAGGAGCTGCGCGCAATCTTCACTGAGGCGAATTTGCGGCGCGCAACGTTGCGT 540
QY 541 CCGGATACGATTCGCAAGTTCTTGTTCATTGCGCTTCAAACTGCTTAAGCGGTG 600
DB 541 CCGGATACGATTCGCAAGTTCTTGTTCATTGCGCTTCAAACTGCTTAAGCGGTG 600
QY 601 CCGGATACGATTCGCAAGTTCTTGTTCATTGCGCTTCAAACTGCTTAAGCGGTG 660
DB 601 CCGGATACGATTCGCAAGTTCTTGTTCATTGCGCTTCAAACTGCTTAAGCGGTG 660
QY 661 TTTGTGTAAGAGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
DB 661 TTTGTGTAAGAGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
QY 721 AGGCAACATTCGCAATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
DB 721 AGGCAACATTCGCAATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780

DB 721 AGGCAACATTCGCAATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
QY 781 CCAACGCGCCCAAGGTTCCGCGAGAGCTTCCGCACTTGAGCGAAATCTTCCCATGCG 840
DB 781 CCAACGCGCCCAAGGTTCCGCGAGAGCTTCCGCACTTGAGCGAAATCTTCCCATGCG 840
QY 841 TACCTCACTGCTGCGAAAGGCTGGAAGAAATGAGTGGTCCCTTGAGCGAGACATACC 900
DB 841 TACCTCACTGCTGCGAAAGGCTGGAAGAAATGAGTGGTCCCTTGAGCGAGACATACC 900
QY 901 CTGCGGGAACGCTTCTTCCGCTGCGCAATGAGCTTCTTCTTCCGCGGCTGCTGCTG 960
DB 901 CTGCGGGAACGCTTCTTCCGCTGCGCAATGAGCTTCTTCTTCCGCGGCTGCTGCTG 960
QY 961 CAAGGATCTGGAATGCTTGAACCGGCTGCTGCAACAGCACTGCTGCTGCTGCTGCT 1020
DB 961 CAAGGATCTGGAATGCTTGAACCGGCTGCTGCAACAGCACTGCTGCTGCTGCTGCT 1020
QY 1021 ATGATGCGGCTGCTGCGATGAGCGAGACTGCTTCTTCTGCACTTTTACCCAGGACCG 1080
DB 1021 ATGATGCGGCTGCTGCGATGAGCGAGACTGCTTCTTCTGCACTTTTACCCAGGACCG 1080
QY 1081 CTGTCGATGCTGCTTACATTTGGGCTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCT 1140
DB 1081 CTGTCGATGCTGCTTACATTTGGGCTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCT 1140
QY 1141 GTGATGGAATTTGAAGGCGCTTTCATGTCGCGCACTGCTGCTGCTGCTGCTGCTGCT 1200
DB 1141 GTGATGGAATTTGAAGGCGCTTTCATGTCGCGCACTGCTGCTGCTGCTGCTGCTGCT 1200
QY 1201 GCTCTGGAACAAATGCGCGCAAGCTTTCATGTCGCGCACTGCTGCTGCTGCTGCTGCT 1260
DB 1201 GCTCTGGAACAAATGCGCGCAAGCTTTCATGTCGCGCACTGCTGCTGCTGCTGCTGCT 1260
QY 1261 ATCAATGCGCAATCTGCGCATCTCAGAAAGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
DB 1261 ATCAATGCGCAATCTGCGCATCTCAGAAAGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
QY 1321 GAAGACTTCAAGCTCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCA 1380
DB 1321 GAAGACTTCAAGCTCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCA 1380
QY 1381 GTTCTGGAAGCGGCTCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCA 1440
DB 1381 GTTCTGGAAGCGGCTCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCA 1440
QY 1441 CTTGSAATGCTGCTGCTTCCGCGCTTCTGCACTGCGCGCTTCTGCGGCGCTGAGAA 1500
DB 1441 CTTGSAATGCTGCTGCTTCCGCGCTTCTGCACTGCGCGCTTCTGCGGCGCTGAGAA 1500
QY 1501 GAGGCGTCCGAGCGCGAGGCTTTCGCAAGTGAACCGGCTTCCGCTGCTGCTGCTGCTG 1560
DB 1501 GAGGCGTCCGAGCGCGAGGCTTTCGCAAGTGAACCGGCTTCCGCTGCTGCTGCTGCTG 1560
QY 1561 CTGAAAGCACTCAATGAGAGCACTGCGAATGCGAATGCGAATGCGAATGCGAATGCG 1620
DB 1561 CTGAAAGCACTCAATGAGAGCACTGCGAATGCGAATGCGAATGCGAATGCGAATGCG 1620
QY 1621 CTGGAATGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1680
DB 1621 CTGGAATGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1680
QY 1681 CCGGCTGCTTTCGCAATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1740
DB 1681 CCGGCTGCTTTCGCAATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1740
QY 1741 TCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1800
DB 1741 TCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1800

RESULT 2
US-09-750-986d-29

Sequence 29. Application US/09750986D
Publication NO. US20030228670A1
GENERAL INFORMATION:
APPLICANT: Steinbuechel, Alexander
APPLICANT: Priefert, Horst
APPLICANT: Rabenhorst, Jurgen
TITLE OF INVENTION: SYNTHETIC ENZYMES FOR THE PRODUCTION OF CONIFERYL
TITLE OF INVENTION: ALCOHOL, CONIFERYLALDEHYDE, FERULIC ACID, VANILLIN AND
TITLE OF INVENTION: VANILLIC ACID AND THEIR USE
FILE REFERENCE: Bayer-9998-CAO
CURRENT APPLICATION NUMBER: US/09/750, 986D
CURRENT FILING DATE: 2000-12-28
PRIOR APPLICATION NUMBER: 196 49 655.1 GERMANY
PRIOR FILING DATE: 1996-11-29
NUMBER OF SEQ ID NOS: 45
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 29
LENGTH: 1770
TYPE: DNA
ORGANISM: Pseudomonas sp.
FEATURE:
NAME/KEY: CDS
LOCATION: (1..(1767))
OTHER INFORMATION: product = "Ferulasaure-CoA-Synthetase" / gene =
OTHER INFORMATION: "fcs"
US-09-750-986D-29

Query Match 100.0%; Score 1770; DB 10; Length 1770;
Best local Similarity 100.0%; Pred. No. 0;
Matches 1770; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCTCTCTCGAGGCGCTTCTCCCTCCGGGTGGAATTCGTGAGCGTCTGAGAT 60
DB 1 ATGGCTCTCTCGAGGCGCTTCTCCCTCCGGGTGGAATTCGTGAGCGTCTGAGAT 60
QY 61 TGGGCTAAGACCCCTCCAGAACAAACCTGCTGCTGCGACGCGGCAATGGGAAATGG 120
DB 61 TGGGCTAAGACCCCTCCAGAACAAACCTGCTGCTGCGACGCGGCAATGGGAAATGG 120
QY 121 CGTGTATACGACTACGCGGAAATGTTCCAAACGTCGCGGCATCGACAGAGCTTGCTT 180
DB 121 CGTGTATACGACTACGCGGAAATGTTCCAAACGTCGCGGCATCGACAGAGCTTGCTT 180
QY 121 CGTGTATACGACTACGCGGAAATGTTCCAAACGTCGCGGCATCGACAGAGCTTGCTT 180
DB 121 CGTGTATACGACTACGCGGAAATGTTCCAAACGTCGCGGCATCGACAGAGCTTGCTT 180
QY 181 CCTTACGCACTATGCGAGACGCTCGCTGCTTATCGTCTTGAAATGAACTGGAACT 240
DB 181 CCTTACGCACTATGCGAGACGCTCGCTGCTTATCGTCTTGAAATGAACTGGAACT 240
QY 181 CCTTACGCACTATGCGAGACGCTCGCTGCTTATCGTCTTGAAATGAACTGGAACT 240
DB 181 CCTTACGCACTATGCGAGACGCTCGCTGCTTATCGTCTTGAAATGAACTGGAACT 240
QY 241 CTTGAGCTGGCAATTTGGGGCTATATGCGGGCAATCCCTATTGCCGGTGTCTCTGCT 300
DB 241 CTTGAGCTGGCAATTTGGGGCTATATGCGGGCAATCCCTATTGCCGGTGTCTCTGCT 300
QY 241 CTTGAGCTGGCAATTTGGGGCTATATGCGGGCAATCCCTATTGCCGGTGTCTCTGCT 300
DB 241 CTTGAGCTGGCAATTTGGGGCTATATGCGGGCAATCCCTATTGCCGGTGTCTCTGCT 300
QY 301 TATTCAGCTGCTGTGCAAGATTTGGCGAAGCTGCGTCATCGTAGGTCTTTCTCAACCG 360
DB 301 TATTCAGCTGCTGTGCAAGATTTGGCGAAGCTGCGTCATCGTAGGTCTTTCTCAACCG 360
QY 301 TATTCAGCTGCTGTGCAAGATTTGGCGAAGCTGCGTCATCGTAGGTCTTTCTCAACCG 360
DB 301 TATTCAGCTGCTGTGCAAGATTTGGCGAAGCTGCGTCATCGTAGGTCTTTCTCAACCG 360
QY 361 GAGCTGCTTTTGTGCTGCGATGACGACCTTTCCAGCGCGCAATTTGAGAACCAATTCGCG 420
DB 361 GAGCTGCTTTTGTGCTGCGATGACGACCTTTCCAGCGCGCAATTTGAGAACCAATTCGCG 420
QY 361 GAGCTGCTTTTGTGCTGCGATGACGACCTTTCCAGCGCGCAATTTGAGAACCAATTCGCG 420
DB 361 GAGCTGCTTTTGTGCTGCGATGACGACCTTTCCAGCGCGCAATTTGAGAACCAATTCGCG 420
QY 421 GAGCAGCTGCGCGCAATCTTCACTCGAGCGCAATTTGGCGGGCGCGCAAGGTATTTT 480
DB 421 GAGCAGCTGCGCGCAATCTTCACTCGAGCGCAATTTGGCGGGCGCGCAAGGTATTTT 480
QY 421 GAGCAGCTGCGCGCAATCTTCACTCGAGCGCAATTTGGCGGGCGCGCAAGGTATTTT 480
DB 421 GAGCAGCTGCGCGCAATCTTCACTCGAGCGCAATTTGGCGGGCGCGCAAGGTATTTT 480
QY 481 GACAGCTGCTGTGAGCAGAGCTGTGGATTGAGGCAAGTAATGCTTTTGGCGCAACTGGC 540
DB 481 GACAGCTGCTGTGAGCAGAGCTGTGGATTGAGGCAAGTAATGCTTTTGGCGCAACTGGC 540
QY 481 GACAGCTGCTGTGAGCAGAGCTGTGGATTGAGGCAAGTAATGCTTTTGGCGCAACTGGC 540
DB 481 GACAGCTGCTGTGAGCAGAGCTGTGGATTGAGGCAAGTAATGCTTTTGGCGCAACTGGC 540
QY 541 CCGCATACGATTGCAAGTCTTGTTCATCTTGGCTTACCAAACTGCTTAAAGCGGTG 600
DB 541 CCGCATACGATTGCAAGTCTTGTTCATCTTGGCTTACCAAACTGCTTAAAGCGGTG 600
QY 541 CCGCATACGATTGCAAGTCTTGTTCATCTTGGCTTACCAAACTGCTTAAAGCGGTG 600
DB 541 CCGCATACGATTGCAAGTCTTGTTCATCTTGGCTTACCAAACTGCTTAAAGCGGTG 600
QY 601 CCGACTACTAGCAATGCTCTGCGCCAAATCAGAGATGCTTTCGAACTTTCCGGTT 660
DB 601 CCGACTACTAGCAATGCTCTGCGCCAAATCAGAGATGCTTTCGAACTTTCCGGTT 660

DB 601 CCGACTACTAGCAATGCTCTGCGCCAAATCAGAGATGCTTTCGAACTTTCCGGTT 660
QY 661 TTTGGTGAAGACCGCGGCTGTGTGATGCTGTGCTGCGGAAACCACTTTCGGGGC 720
DB 661 TTTGGTGAAGACCGCGGCTGTGTGATGCTGTGCTGCGGAAACCACTTTCGGGGC 720
QY 721 AGCCAAACATCGGATCGTGTGTGCAACGCGCGGCACTACTACTGACGACGTTAA 780
DB 721 AGCCAAACATCGGATCGTGTGTGCAACGCGCGGCACTACTACTGACGACGTTAA 780
QY 781 CCAACCGCCCAAGGTTTCGCGAGACGCTTTCGCAACTTGAACGAAATCTTCCACTGCG 840
DB 781 CCAACCGCCCAAGGTTTCGCGAGACGCTTTCGCAACTTGAACGAAATCTTCCACTGCG 840
QY 841 TACCTCACTGTGCGGAAAGCTGTGGAAGAAATTAAGTGGTGCCTTTGAGCCGAGATACC 900
DB 841 TACCTCACTGTGCGGAAAGCTGTGGAAGAAATTAAGTGGTGCCTTTGAGCCGAGATACC 900
QY 901 CTTGCGCAACGCTTCTTCGCTGCGATGAGCTGTTCTTCTTCGCGGCGGCTGGTTGTCG 960
DB 901 CTTGCGCAACGCTTCTTCGCTGCGATGAGCTGTTCTTCTTCGCGGCGGCTGGTTGTCG 960
QY 961 CAAAGGATCTGGATCGTTTGAACCGGCTCGCTGAACAGCACTGTGTGAGCGCAATTGCG 1020
DB 961 CAAAGGATCTGGATCGTTTGAACCGGCTCGCTGAACAGCACTGTGTGAGCGCAATTGCG 1020
QY 1021 ATGATGCGGGTCTGGGCAATGACGGAATGCTCTTCTTCGCACTTTTACACCGGACCG 1080
DB 1021 ATGATGCGGGTCTGGGCAATGACGGAATGCTCTTCTTCGCACTTTTACACCGGACCG 1080
QY 1081 CTTGCAATGCTGTGTTACATTTGAGGCTGCGACGCGCTGCTGCGAGCTCAAGCTGTTCCG 1140
DB 1081 CTTGCAATGCTGTGTTACATTTGAGGCTGCGACGCGCTGCTGCGAGCTCAAGCTGTTCCG 1140
QY 1141 GTGATGGAATTTGAAGGCGCTTTCATGCTCGGACGTCATGAGCGGCTACTGCGCT 1200
DB 1141 GTGATGGAATTTGAAGGCGCTTTCATGCTCGGACGTCATGAGCGGCTACTGCGCT 1200
QY 1201 GCTTCTGAACAAATGCGCCAAAGCTTTCGACGAGGAAGGCTTATCTGCTCCGGTGAATGCC 1260
DB 1201 GCTTCTGAACAAATGCGCCAAAGCTTTCGACGAGGAAGGCTTATCTGCTCCGGTGAATGCC 1260
QY 1261 ATCAAAATGGAAGATCTCGCGATCTCAAGAAAGGCTGATGTTTGAACGCTGCAATTGCT 1320
DB 1261 ATCAAAATGGAAGATCTCGCGATCTCAAGAAAGGCTGATGTTTGAACGCTGCAATTGCT 1320
QY 1321 GAAGACTTCAAGCTGCTCAAGGAGTATTTGTACGCTGTGGCCATTGCGCACGCGGCG 1380
DB 1321 GAAGACTTCAAGCTGCTCAAGGAGTATTTGTACGCTGTGGCCATTGCGCACGCGGCG 1380
QY 1381 GTTCTGGAAGCGGCTCTTACGCTCTGAGAGTATGTTGTGCTGCTTGAATCGGAATGC 1440
DB 1381 GTTCTGGAAGCGGCTCTTACGCTCTGAGAGTATGTTGTGCTGCTTGAATCGGAATGC 1440
QY 1441 CTTGATGCTGCTGCTTCCGCGCTTCTGCACTGCGGCTGCTTGTGCGGGCTAGGAAA 1500
DB 1441 CTTGATGCTGCTGCTTCCGCGCTTCTGCACTGCGGCTGCTTGTGCGGGCTAGGAAA 1500
QY 1501 GAGGCTGCGACGCGGAGTCTTTCAGATGAGCGGCTGTGGGCTGTGACTGCG 1560
DB 1501 GAGGCTGCGACGCGGAGTCTTTCAGATGAGCGGCTGTGGGCTGTGACTGCG 1560
QY 1561 CTCAAAAGATCTCAATCGAAGAACCACTGCGCAATCCAGTGGCATATGTTGGGTGCGTC 1620
DB 1561 CTCAAAAGATCTCAATCGAAGAACCACTGCGCAATCCAGTGGCATATGTTGGGTGCGTC 1620
QY 1621 CTTGATGCGCGCGCTGATGATTAAGGCGAGGCTCACTGACAAAGGCTCGATCAACAG 1680
DB 1621 CTTGATGCGCGCGCTGATGATTAAGGCGAGGCTCACTGACAAAGGCTCGATCAACAG 1680
QY 1681 CGGCTGTTTTCGAATGCGGCTGCGGCAAGTTGATGCGCTGATGCTGTGAAGATCAA 1740
DB 1681 CGGCTGTTTTCGAATGCGGCTGCGGCAAGTTGATGCGCTGATGCTGTGAAGATCAA 1740

Qy 1741 TCCATGCTGCTGACGAGCCACACTGTGA 1770
Db 1741 TCCATGCTGCTGACGAGCCACACTGTGA 1770

RESULT 3

US-08-976-063C-1
Sequence 1, Application US/08976063C
Publication No. US20020182697A1

GENERAL INFORMATION:

APPLICANT: Alexander Steinduchel; Horst Priefert; Jurgen Rabenhorst

TITLE OF INVENTION: SYNTHETIC ENZYMES FOR THE PRODUCTION OF

TITLE OF INVENTION: CONFERYL ALCOHOL, CONFERYLALDEHYDE, FERULIC ACID, VANILLIN A

TITLE OF INVENTION: ACID AND THEIR USE

NUMBER OF SEQUENCES: 42

CORRESPONDENCE ADDRESSES:

ADDRESSEE: SPRUNG KRAMER SCHAEFER & BRISCOE

STREET: 660 White Plains Road

CITY: Tarrytown

STATE: New York

COUNTRY: U.S.A.

ZIP: 10591-5144

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB storage

COMPUTER: HP VECTRA

OPERATING SYSTEM: DOS

SOFTWARE: Wordperfect 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/976, 063C

FILING DATE: 21-NOV-1997

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 196 49 655.1 (Germany)

FILING DATE: 29-NOV-1996

ATTORNEY/AGENT INFORMATION:

NAME: Kurt G. Briscoe

REGISTRATION NUMBER: 33,141

REFERENCE/DOCKET NUMBER: Bayer 9998-CAO

TELECOMMUNICATION INFORMATION:

TELEPHONE: (914) 332-1700

TELEFAX: (914) 332-1844

TELEX:

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 32679 base pairs

TYPE: Nucleic Acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Pseudomonas sp.

STRAIN: HR199

FEATURE:

NAME/KEY: CDS

LOCATION: 3146..3997

OTHER INFORMATION: /gene= "ORF1"

US-08-976-063C-1

Query Match 100.0%; Score 1770; DB 8; Length 32679;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1770; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGCGTCTCTCGAGGCGCTTCTTCCCTGCGGCTGGAATTCTTGAGCGTCTCGACAT 60

Db 21288 ATGCGTCTCTCGAGGCGCTTCTTCCCTGCGGCTGGAATTCTTGAGCGTCTCGACAT 21347

Qy 61 TGGGCTAAGACCCGTCAGAAACAACCTGCGTTCGCGAGGGGCGCAAAATGGGGAATGG 120

Db 21348 TGGGCTAAGACCCGTCAGAAACAACCTGCGTTCGCGAGGGGCGCAAAATGGGGAATGG 21407

Qy 121 CGTCGATCAGCTACGCGGAATGTTCCACAACGTCGCGCCATCGCACAGACTTGCTT 180

Db 21408 CGTCGATCAGCTACGCGGAATGTTCCACAACGTCGCGCCATCGCACAGACTTGCTT 21467

Qy 181 CTTTACGGACTATCGGACAGAGCGTCCGCTGCTTATGCTCTCTGGAATATGACTTGAACAT 240

Db 21468 CTTTACGGACTATCGGACAGAGCGTCCGCTGCTTATGCTCTCTGGAATATGACTTGAACAT 21527

Qy 241 CTTACGCTGCAATTTGGGGCTATGTAAGCGGGGATTCCTTATTTGCCCGGTCTCTGCT 300

Db 21528 CTTACGCTGCAATTTGGGGCTATGTAAGCGGGGATTCCTTATTTGCCCGGTCTCTGCT 21587

Qy 301 TATTCATGCTGTCGCAAGATTTGGCGAAGCTGCTCATCTGATGCTTCTTGCACCG 360

Db 21588 TATTCATGCTGTCGCAAGATTTGGCGAAGCTGCTCATCTGATGCTTCTTGCACCG 21647

Qy 361 GGAATGCTCTTTCCTGCTCCATGAGACACCTTTCACAGCGGCAATTGACATTTGCGCG 420

Db 21648 GGAATGCTCTTTCCTGCTCCATGAGACACCTTTCACAGCGGCAATTGACATTTGCGCG 21707

Qy 421 GACGAGCTGCGCCCAATCTTCACTCGAGGCGAATTGAGCGGCGGCGCACGCTGAGTTT 480

Db 21708 GACGAGCTGCGCCCAATCTTCACTCGAGGCGAATTGAGCGGCGGCGCACGCTGAGTTT 21767

Qy 481 GACAGCTGCTGAGCAGCGCTGAGGATTTAGGACAGATTAATGCTTTGGGCACTGGC 540

Db 21768 GACAGCTGCTGAGCAGCGCTGAGGATTTAGGACAGATTAATGCTTTGGGCACTGGC 21827

Qy 541 CCGGATACGATTTGCCAAGTTCTTGTTCATCTTGGCTCTTACCAAACTGCTTAAAGCGGTG 600

Db 21828 CCGGATACGATTTGCCAAGTTCTTGTTCATCTTGGCTCTTACCAAACTGCTTAAAGCGGTG 21887

Qy 601 CCGACTACTGAGGGAATGCTCTGCGCAATCAGACAGATGCTTCTGCAAACTTTCCGCTT 660

Db 21888 CCGACTACTGAGGGAATGCTCTGCGCAATCAGACAGATGCTTCTGCAAACTTTCCGCTT 21947

Qy 661 TTTGTGGAAGAGCGCGGCTGCTGAGTGTGAGTGTGCGTGGCAACACCTTCCGCGGC 720

Db 21948 TTTGTGGAAGAGCGCGGCTGCTGAGTGTGAGTGTGCGTGGCAACACCTTCCGCGGC 22007

Qy 721 AGCCACAACATCGGCAATCGTGTGTAACAAGGGGCAAGATTAATCTTGAAGCGGTAAG 780

Db 22008 AGCCACAACATCGGCAATCGTGTGTAACAAGGGGCAAGATTAATCTTGAAGCGGTAAG 22067

Qy 781 CCAAGCGCCCAAGGTTGCGCGAGAGCTTTCGCACTTGAAGCGAAATCTTCCCACTGCG 840

Db 22068 CCAAGCGCCCAAGGTTGCGCGAGAGCTTTCGCACTTGAAGCGAAATCTTCCCACTGCG 22127

Qy 841 TACCTCACTGTGCGAAGGCTGGAAGAAATTAATGAGGTCCTTGAAGCAGACAGTACC 900

Db 22128 TACCTCACTGTGCGAAGGCTGGAAGAAATTAATGAGGTCCTTGAAGCAGACAGTACC 22187

Qy 901 CTGCGGGAACGCTTCTGCTGCGATGAAGCTGTTCTTCTGCGGCGGCTGAGGTTGTG 960

Db 22188 CTGCGGGAACGCTTCTGCTGCGATGAAGCTGTTCTTCTGCGGCGGCTGAGGTTGTG 22247

Qy 961 CAAGGATCTGAGATGTTTGAACCGGCTGCTGGAACAGCACTGTGAGTGAAGCCATTGCG 1020

Db 22248 CAAGGATCTGAGATGTTTGAACCGGCTGCTGGAACAGCACTGTGAGTGAAGCCATTGCG 22307

Qy 1021 ATGATGCGGGGTCTGAGCATGACGAGACTGCTCTTCTGCACTTTTACACCGGACCG 1080

Db 22308 ATGATGCGGGGTCTGAGCATGACGAGACTGCTCTTCTGCACTTTTACACCGGACCG 22367

Qy 1081 CTGTGATGCTGCTTACATTTGGGCTGCAAGCGGCTGCGCGAGGCTCAAGCTCGTTCCG 1140

Db 22368 CTGTGATGCTGCTTACATTTGGGCTGCAAGCGGCTGCGCGAGGCTCAAGCTCGTTCCG 22427

Qy 1141 GTCGATGGAATTTGAAGGCGCTTTCATATGTCGCCACGTCATGAGCGGCTACTGGCGT 1200

Db 22428 GTCGATGGAATTTGAAGGCGCTTTCATATGTCGCCACGTCATGAGCGGCTACTGGCGT 22487

Qy 1201 GCTCTGAACAAATAATGCCCAAGCTTGCAGAGGAAGGCTATTATGCTCCGCTGATGCC 1260

Db 22488 GCTCTGGAACAAATGCTCCAAAGGTTGAGAGAGAAAGCTATTACTCTCCGGTGAAGGCC 22547
Qy 1261 ATCAAAATGGCAGATCTCTCCGATCTCTAGAAAAGCTTGATGTTTGAAGCTCGAATTGCT 1320
Db 22548 ATCAAAATGGCAGATCTCTCCGATCTCTAGAAAAGCTTGATGTTTGAAGCTCGAATTGCT 22607
Qy 1321 GAAGACTTCAAGTGTCTCAAGGGGATATTGTACAGCGCTTGGGCAATGGCCAGCGGGGCG 1380
Db 22608 GAAGACTTCAAGTGTCTCAAGGGGATATTGTACAGCGCTTGGGCAATGGCCAGCGGGGCG 22667
Qy 1381 GTTCTGAAGAGCGGCTCTTAAGTCTCTGAGAGTGTGCTGCTCTCTGATTCGTAATGC 1440
Db 22668 GTTCTGAAGAGCGGCTCTTAAGTCTCTGAGAGTGTGCTGCTCTCTGATTCGTAATGC 22727
Qy 1441 CTTCGATGCTGCTGTTTCCGCGCTTCTCTGACTGCGCTGCTTGTCTGGGAGTAAGAAA 1500
Db 22728 CTTCGATGCTGCTGTTTCCGCGCTTCTCTGACTGCGCTGCTTGTCTGGGAGTAAGAAA 22787
Qy 1501 GAGGCGTCCGAGCGCCGAGGCTTTCAGTGAAGCGGCTTGGGCTGCTGCTGCTGACTGG 1560
Db 22788 GAGGCGTCCGAGCGCCGAGGCTTTCAGTGAAGCGGCTTGGGCTGCTGCTGCTGACTGG 22847
Qy 1561 CTCGAAAGACTCAATGAGAGCACTGAGCAATGCGCATGCTGAGGAGGCTC 1620
Db 22848 CTCGAAAGACTCAATGAGAGCACTGAGCAATGCGCATGCTGAGGAGGCTC 22907
Qy 1621 CTCGATACCGCCGCTGCTGATTAAGAGGCGAGGCTCACTGACAAAGGCTGATCAACAG 1680
Db 22908 CTCGATACCGCCGCTGCTGATTAAGAGGCGAGGCTCACTGACAAAGGCTGATCAACAG 22967
Qy 1681 CGGCTGTTTTCGATAGGCGGCTGCGGAGAAAGTTGATGCGCTGATCGTGGTAAGATCAA 1740
Db 22968 CGGCTGTTTTCGATAGGCGGCTGCGGAGAAAGTTGATGCGCTGATCGTGGTAAGATCAA 23027
Qy 1741 TCCATGCTGCTGAGCAGAGCCACACTGTGA 1770
Db 23028 TCCATGCTGCTGAGCAGAGCCACACTGTGA 23057

RESULT 4
US-09-750-986d-1
; Sequence 1, Application US/09750986D
; Publication No. US20030228670A1
; GENERAL INFORMATION:
; APPLICANT: Steinduchel, Alexander
; APPLICANT: Pfeleert, Horst
; APPLICANT: Rabenhorst, Jurgen
; TITLE OF INVENTION: SYNTHETIC ENZYMES FOR THE PRODUCTION OF CONIFERYL
; TITLE OF INVENTION: ALCOHOL, CONFERYLALDEHYDE, FERULIC ACID, VANILLIN AND
; FILE REFERENCE: VANILLIC ACID AND THEIR USE
; CURRENT APPLICATION NUMBER: US/09/750, 986D
; PRIOR FILING DATE: 2000-12-28
; PRIOR FILING DATE: 1996-11-29
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 32679
; TYPE: DNA
; ORGANISM: Pseudomonas sp.
; NAME/KEY: CDS
; LOCATION: (3146)..(3997)
; OTHER INFORMATION: gene = "ORF1"
US-09-750-986d-1

Query Match 100.0%; Score 1770; DB 10; Length 32679;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1770; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGCGTCTCTCAGAGCGCTTCTCCCTCCGGGTCGAATTCTTGAGCGCTTCGAGCAT 60

Db 21288 ATGCGTCTCTCAGAGCGCTTCTCCCTCCGGGTCGAATTCTTGAGCGCTTCGAGCAT 21347
Qy 61 TGGGCTTAAGACCCGCTCCAGAACAAACTGCTGTGCTGCCAGGGCGGCAATGGGGAATGG 120
Db 21348 TGGGCTTAAGACCCGCTCCAGAACAAACTGCTGTGCTGCCAGGGCGGCAATGGGGAATGG 21407
Qy 121 CGTGTATCAGTACGCGGGAATGTTCCACAAGCTCCGGGCAATGCGACAGAGCTTGGCT 180
Db 21408 CGTGTATCAGTACGCGGGAATGTTCCACAAGCTCCGGGCAATGCGACAGAGCTTGGCT 21467
Qy 181 CCTTAAGCAATATGCGAGAGCGCTCCGCTGCTTATGCTCTGGAATGAAGCTGGAACAT 240
Db 21468 CCTTAAGCAATATGCGAGAGCGCTCCGCTGCTTATGCTCTGGAATGAAGCTGGAACAT 21527
Qy 241 CTTCAGCTGCAATTTGGGCGTATGATATGCGGCGCAATTCCTATTGCCGCGTCTCTGCT 300
Db 21528 CTTCAGCTGCAATTTGGGCGTATGATATGCGGCGCAATTCCTATTGCCGCGTCTCTGCT 21587
Qy 301 TATTCACTGCTGTGCGAAGATTGGGGAAGCTGTCACATCGTAGGCTTCTGCAACCG 360
Db 21588 TATTCACTGCTGTGCGAAGATTGGGGAAGCTGTCACATCGTAGGCTTCTGCAACCG 21647
Qy 361 GGACTGCTTTCGCTGCGAGTGCAGACCTTTCAGCGCGCAATTGAGCAATTCTGCGC 420
Db 21648 GGACTGCTTTCGCTGCGAGTGCAGACCTTTCAGCGCGCAATTGAGCAATTCTGCGC 21707
Qy 421 GAGCAGCTGCGCGCAATCTTCACTCGAGGCGAATTGGCGGCGGCGGCGAGGCTGATTT 480
Db 21708 GAGCAGCTGCGCGCAATCTTCACTCGAGGCGAATTGGCGGCGGCGGCGAGGCTGATTT 21767
Qy 481 GAGCAGCTGCTGAGAGAGCTGCTGAGGATTGAGGAGATTAATGCTTGGCGCAACTGGC 540
Db 21768 GAGCAGCTGCTGAGAGAGCTGCTGAGGATTGAGGAGATTAATGCTTGGCGCAACTGGC 21827
Qy 541 CCGGATACGATTGCGCAAGTCTTGTTCATCTTGGCTCTACCAACTGCTTAAGCGCGTG 600
Db 21828 CCGGATACGATTGCGCAAGTCTTGTTCATCTTGGCTCTACCAACTGCTTAAGCGCGTG 21887
Qy 601 CCGACTTACTAGCGAATGCTCTGCGGCAATCAGCAGATGCTTCTGCAAACTTTCCGCGTT 660
Db 21888 CCGACTTACTAGCGAATGCTCTGCGGCAATCAGCAGATGCTTCTGCAAACTTTCCGCGTT 21947
Qy 661 TTTGGTGAAGAGCGCGGCTGCTGAGTGTGAGTGTGGTGGCGGGAACACACTTCCGCGCG 720
Db 21948 TTTGGTGAAGAGCGCGGCTGCTGAGTGTGAGTGTGGTGGCGGGAACACACTTCCGCGCG 22007
Qy 721 AGCCACACATCCGCGCATCTGTTGTACACGCGGCGAGCTACTTACCTTGAAGCGGTAA 780
Db 22008 AGCCACACATCCGCGCATCTGTTGTACACGCGGCGAGCTACTTACCTTGAAGCGGTAA 22067
Qy 781 CCAACCGCCCAAGGGTTCGCGGAGAGCTTTCGCACTTGAGCGAATCTTCCACTGCG 840
Db 22068 CCAACCGCCCAAGGGTTCGCGGAGAGCTTTCGCACTTGAGCGAATCTTCCACTGCG 22127
Qy 841 TACCTACTGTGCGAAGGCTGCGGAGGATTAATGCTGCTCCCTTGAAGGAGACAGTACC 900
Db 22128 TACCTACTGTGCGAAGGCTGCGGAGGATTAATGCTGCTCCCTTGAAGGAGACAGTACC 22187
Qy 901 CTGCGGGAACGCTTCTTCCGCTCGCATGAAGCTTCTTCTCCGCGGCGCTGCGGTGTCG 960
Db 22188 CTGCGGGAACGCTTCTTCCGCTCGCATGAAGCTTCTTCTCCGCGGCGCTGCGGTGTCG 22247
Qy 961 CAAGGATCTGGGATGTTTGAACCGGCTGCTGAACAGCATGCTGTGAGCGCATTTGCG 1020
Db 22248 CAAGGATCTGGGATGTTTGAACCGGCTGCTGAACAGCATGCTGTGAGCGCATTTGCG 22307
Qy 1021 ATGATGCGGCGGCTGAGGATGAGGAGACTGCTTCTTCCGCACTTTTACACGAGACCG 1080
Db 22308 ATGATGCGGCGGCTGAGGATGAGGAGACTGCTTCTTCCGCACTTTTACACGAGACCG 22367
Qy 1081 CTGTCATGAGCTGCTGATTAATTTGGGCTGCGAGCGCTGCTGCGAGGTCAGAGCTGTTCCG 1140

Db 22368 CTGTCAGATGCTGCTTACATTGGGCTGCCAGCCCTGCTGCGAGGTCAGTCTGTTCCG 22427
Qy 1141 GTTCGATGGGAAATTTGAAAGGGGCTTTCATGTCGCGACGTCATGAGCGGCTACTGCGCG 1200
Db 22428 GTTCGATGGGAAATTTGAAAGGGGCTTTCATGTCGCGACGTCATGAGCGGCTACTGCGCG 22487
Qy 1201 GCTCTGAAACAAATATCCCAAGGCTTTCAGAGGAGGCTATTACTCTCCGCTGATGCG 1260
Db 22488 GCTCTGAAACAAATATCCCAAGGCTTTCAGAGGAGGCTATTACTCTCCGCTGATGCG 22547
Qy 1261 ATCAAAATTTGGCAGATCTCTCCGATCTCTCAGAAAGCTGATGTTTGAACGCTGCAATTTGCT 1320
Db 22548 ATCAAAATTTGGCAGATCTCTCCGATCTCTCAGAAAGCTGATGTTTGAACGCTGCAATTTGCT 22607
Qy 1321 GAAGACTTCAAGCTCTCTCAGGAGGATTTGTACAGGCTTGGGCAATTTGCGACGCGGCG 1380
Db 22608 GAAGACTTCAAGCTCTCTCAGGAGGATTTGTACAGGCTTGGGCAATTTGCGACGCGGCG 22667
Qy 1381 GTTCTGGAAGCGGCTCTTACGTCCTGAGAGTGTGCTGCTCCTGATCTGTAATGC 1440
Db 22668 GTTCTGGAAGCGGCTCTTACGTCCTGAGAGTGTGCTGCTCCTGATCTGTAATGC 22727
Qy 1441 CTTCGATGCTGCTGCTTTTCCGCTCTTCTGACTGCGGCTCTTGTGCGGCTAGAGAAA 1500
Db 22728 CTTCGATGCTGCTGCTTTTCCGCTCTTCTGACTGCGGCTCTTGTGCGGCTAGAGAAA 22787
Qy 1501 GAGGCGTCCGAGCGCGAGGCTTTCAGAGTGTGCTGCGGCTCTGCTGCTGCTGCTGCTG 1560
Db 22788 GAGGCGTCCGAGCGCGAGGCTTTCAGAGTGTGCTGCGGCTCTGCTGCTGCTGCTGCTG 22847
Qy 1561 CTCGAAAGCATCTCAATCGAAGAGCACTGGCAATGCCAGTGCATCATGTGGTAGGGCTC 1620
Db 22848 CTCGAAAGCATCTCAATCGAAGAGCACTGGCAATGCCAGTGCATCATGTGGTAGGGCTC 22907
Qy 1621 CTCGATACGCGCGCGCTGCTGATGATTAAGGGGAGAGTCACTGACAAAGGCTCTGATCAACCG 1680
Db 22908 CTCGATACGCGCGCGCTGCTGATGATTAAGGGGAGAGTCACTGACAAAGGCTCTGATCAACCG 22967
Qy 1681 CGGCGCTTTTTCGATGAGCGGCTGGCGGAAAGTGTGCTGATGCTGATGATGATGATG 1740
Db 22968 CGGCGCTTTTTCGATGAGCGGCTGGCGGAAAGTGTGCTGATGCTGATGATGATGATG 23027
Qy 1741 TCCATGCTGCTGAGCAGAGGCGACACTGTGA 1770
Db 23028 TCCATGCTGCTGAGCAGAGGCGACACTGTGA 23057

RESULT 5
US-10-369-493-44177
; Sequence 44177, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 44177
; LENGTH: 1725
; TYPE: DNA
; ORGANISM: Rhodospseudomonas palustris
US-10-369-493-44177

Query Match 29.8%; Score 527.8; DB 15; Length 1725;
Best Local Similarity 56.6%; Pred. No. 4e-155;

Matches 976; Conservative 0; Mismatches 747; Indels 0; Gaps 0;
Qy 2 TGGGTTCTTCGAGGCGCTTCTCCCTTCGCGGTCGATTTCTTGAAGCTCTGACATTT 61
Db 2 TGGGTTCTTCGAGGCGCTTCTCCCTTCGCGGTCGATTTCTTGAAGCTCTGACATTT 61
Qy 62 GGGCTAAGACCCGCTCAGAAACAACTGCGTTGCTGCGAGGCGGCAATTTGGAAATGCG 121
Db 62 TGGCGAGACCGGCGCCGACCGGCTGTTTATGATGATGATGATGATGATGATGATGATG 121
Qy 122 GTTCGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 181
Db 122 GATGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 181
Qy 182 CTTCAGGATCTATCGGAGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 241
Db 182 CGGCGGATCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 241
Qy 242 TTCAAGCTGATTTGGGCTATGATGATGATGATGATGATGATGATGATGATGATGATG 301
Db 242 CGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 301
Qy 302 ATTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 361
Db 302 ATTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 361
Qy 362 GACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 421
Db 362 GACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 421
Qy 422 ACAGAGTGCAGGCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 481
Db 422 AGGATGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 481
Qy 482 ACAGAGTGCAGGCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 541
Db 482 ACAGAGTGCAGGCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 541
Qy 542 CGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 601
Db 542 CGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 601
Qy 602 CGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 661
Db 602 TCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 661
Qy 662 TGGTGAAGAGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 721
Db 662 TGAAGAGAGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 721
Qy 722 GCGACAAATCGGATCGTGTGTAACAGCGGCGGCACTGTAACCTTGAACGAGTAAAC 781
Db 722 ACCAACAATCGGATCGTGTGTAACAGCGGCGGCACTGTAACCTTGAACGAGTAAAC 781
Qy 782 CAAACGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 841
Db 782 CAAACGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 841
Qy 842 ACTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 901
Db 842 ACTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 901
Qy 902 TGGCGAAGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 961
Db 902 TGGCGAAGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 961
Qy 962 AAGGATCTGAGATGCTTGAACCGGCTGCTGTAACAGCACTGTGTGAGCGCATTTGCA 1021
Db 962 CGCATGCTGGAACCGGCTTGAACCGGCTGCTGTAACAGCACTGTGTGAGCGCATTTGCA 1021
Qy 1022 TGATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1081
Db 1022 TGCTACCGGCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1081

Oy	1082	TCGATGGGAAATTTGAAAGGCGCTTTCCATGATGGTCCGACAGCTCTGAGAGTCAAGCTGTTCCGG	1141
Db	1082	CCAGTCCGCTCCGGCCATGTCGGGCTCCGGGTGCCCGGCACAGAGGCCCAAGCTGGGCGGA	1141
Oy	1142	TCGATGGGAAATTTGAAAGGCGCTTTCCATGATGGTCCGACAGTCAATAGCGGCTACTGGCGCTG	1201
Db	1142	ACAAAGGCAAGCTCGAAGTCCGGGCCAAGGGGCCGAACATCAACCCCGGCTATTGGCGCG	1201
Oy	1202	CTCTCGAACAAAATGCCCAAGCGTTGACAGAGAAAGCTATTACTCTCCGATGATGCCA	1261
Db	1202	CGCCCGAGCTGACCGAATAGGCGCTTCACAGAGGGGCTTCTCAAGCTCAACGATGGGC	1261
Oy	1262	TCAAATTGGCAGATCTCGCGGATCTCAGAAAGGCTCGATGTTTGAACGGTCCGAATGCTG	1321
Db	1262	TCAAAGCCGGTCGATGCGACAGCACTTTTCGGCGGCTTCGATTTGACGGCCGGATCTGG	1321
Oy	1322	AAGACTTCAAGCTGTCTCTCAGGGGATATTTGTACAGCGTTGGCCATTTGGCAACGGCGGGCG	1381
Db	1322	AAGACTTCAAACTGGCGGTGGGGCACTGGGTCACGCTCGGCTACGCCCTACGCCCAAGTTCA	1381
Oy	1382	TTCTGGAAGGCGGCTCTTACGTCCTGGAAGTAGTGTGTGCTGCTCTGATGTGAAATGCC	1441
Db	1382	TTGCCCGCTCGCGCTCTCTGTGTGCGGACAGCTGTGATTCGCGGCTGCACCGCATTAACG	1441
Oy	1442	TTGATTTGCTCGTGTTCGCGCTTCTCTGACTGCGCGTGCCTTTGCGGGGCTAGAGAAAG	1501
Db	1442	TCACCGCGCTGGGAGATCTTCGATTCGCAACGGTGTCAAGATGATCAATCGAACGTCGCCG	1501
Oy	1502	AGGCGTCGGAACGCGCAGAGTCTTGGCAGTAGACCGGTTTCGGGCTGATTTCGTGACTGC	1561
Db	1502	TGGAAAGACTTCGCGCGGCATGTGGCGGCCACCATCTGATTCGGAGAGCGTTTCGCGAGCCCT	1561
Oy	1562	TCAAACGACTCAATCGAAGAACAACTGGGCATGCGATGCGATGATGTGGTAAAGGCTCC	1621
Db	1562	TGCCACGCTGCTGACGACGAGGCCACCGGCTTCGTCACACCGGCTCACCCGCGCTGCTGC	1621
Oy	1622	TCGATACGCGCGCGTCGATTTGATTAAGGCGCAGGTCACATGACAAAGGCTCGATTCACACGC	1681
Db	1622	TCGGGGAACCGCTGTGATCGACAAAGGGGAGATCAACGACAAAGGCTCGGTCACACGAC	1681
Oy	1682	GCGCTGTTTTGCATAGCGCGGTCCGCGCAAGTTGATGGCGCTA	1724
Db	1682	GGGCGGTGCTGAAATACGGCCTCGTTGATCGGAACTTTTA	1724

RESULT 6
 US-10-369-493-35384
 Sequence 35384, Application US/10369493
 Publication No. US2003023675A1
 GENERAL INFORMATION:
 APPLICANT: Cao, Yongwei
 APPLICANT: Hinkle, Gregory J.
 APPLICANT: Slater, Steven C.
 APPLICANT: Goldman, Barry S.
 APPLICANT: Chen, Xianfeng
 TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 TITLES OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 FILE REFERENCE: 38-10(52052)B
 CURRENT APPLICATION NUMBER: US/10/369,493
 CURRENT FILING DATE: 2003-02-28
 PRIOR APPLICATION NUMBER: US 60/360,039
 PRIOR FILING DATE: 2002-02-21
 NUMBER OF SEQ ID NOS: 47374
 SEQ ID NO 35384
 LENGTH: 1728
 TYPE: DNA
 ORGANISM: Agrobacterium tumefaciens
 US-10-369-493-35384

Query Match	22.5%	Score 398.4;	DB 15;	Length 1728;
Best Local Similarity	52.8%	Pred. No. 2.2e-114;		
Matches 908; Conservative	0;	Mismatches 806;	Indels 6;	Gaps 2

Oy	13	GAGGGCGCTTCTTCCTTCCCTCCGGGATCGAATTTCTAGAGGCTCTCGAGCATTTGGGCTAAAGC	72
Db	1	GATCCGCTTGGCCCTTATCCGAGAAATTTAGATGCGCTCTCTTCATCTGGTGTCTTCG	60
Oy	73	CGTCCAGAACAAACCTGCGTTGCTCCGAGGCGGCAATGGGGAATGGCGTGTATCAGC	132
Db	61	GCACCGGAGCGCAATGATGAGCGGACCGTCAAGGGTCCGAGGCATGGCGCGGGTCAAGC	120
Oy	133	TACCGGAAATGTTTCCACACGTCCGGCCATCCGACAGACTTGTCTTACGAACTA	192
Db	121	TACGCGGAGGCTCTCGACAAATACAGGCGTATCGGCGAGTTCCTTCTGATATACCTT	180
Oy	193	TCGGCAGAGCGCCGCTGCTTATGCTCTGAGAAATGACATCTTCAGCTGSCA	252
Db	181	TCGGTCGAGAGGCGCGCTCTGTGTGCTTCTGAAATTCGATCGAGCATGCGTTGATGTTG	240
Oy	253	TTTGGGGCTATGATATGCGGAGCATTCCTTATTTGCCCGGTGTCTCTGTTATTCAGTCTG	312
Db	241	CTTGGCGGCCACAGCATGTGCGTATGCGCTCGGCTGCGCATCAGCGCGGCTTATGCG	297
Oy	313	TCGCAAGATTTGGCGAAGCTGCTACATCGTAGGTCTTTCGCAACCGGGACGTGGTCTTT	372
Db	298	TCGGCGGATCTCACAACTGCGGACATCCGCGGCGAGATCACCCCGGGAGATGGTITTT	357
Oy	373	GCTGCGGATGAGACACTTTCACAGCGCGCAATTTAGACATTTCTGCCGAGCAGGTGCC	432
Db	358	GCCAGGATGCGAGCGCCCTTTGCGCGGGCGCTCGGCGAGGTTTTCGACGACGATACACG	417
Oy	433	GCAATCTTCACTCGAGCGGAATTTGGCCGGGCGGCGC--ACGGTAGTTTTCACAGCTG	489
Db	418	CTGTGTGCGCTCCCGCAACCTGCGGAAAGATGTTTCCAAATAGCTTTCATTTGGAAACCTG	477
Oy	490	CTGAGCAGAGCTGGTGGGATTTGAGGCAATATATGCTTTCGGGCAACTGGCGCCGATAG	549
Db	478	CTTGAAACGAGACCGACCGAAGCGGTTATGTGCTTTCGATGCGGTGCGGCGCGATACC	537
Oy	550	ATTGCGCAAGTTCCTGTTCACTTTCGCTCTACAACTGCTTAAGCGGTGCCGACTACT	609
Db	538	GTCGCGCAATTCCTTTTACGTCAGGCGACGACTGGTTGGCCAAAGCGGTATCCAGACG	597
Oy	610	CAGCGAATGCTCTGCGGCCAATCAGCAGATGCTTCTGCAACTTTCCCGGTTTTGTGTAA	669
Db	598	CAACGCAATGTTGTGTCTCAATCAGAAATGATGGCCGATGTATGCTACTTTTCGCAAG	657
Oy	670	GAGCGCGCGGTGCTGATGGAATGGTGGCCGTGAGAACCACTTCGCGCGGACCGACCAAC	729
Db	658	GAGCGCGCGGTGTGTGTTGATTTGGCGCGCATGGAACCAACGCGCGCGGGAACAAAGTG	717
Oy	730	ATCGGCATGTGTGTATCAACGCGCGGACGTACTACCTTGAAGACGATAAACCAACGCGC	789
Db	718	TTCAACCTGCTCTCTATATAGCGGCTCTCAATTAATATGATTCGCGGCAACCCGAGCCG	777
Oy	790	CAAGGCTTCGCGAGAGCGCTTCGCAACTTGAGCGAAATCTCCACACTGGTAACTCACT	849
Db	778	GCGCAGATGCGCAACACTGACATATCCGAGGACATATGCGCCACTCGGTATTTTCAAC	837
Oy	850	GTCGCGAAGGCTGGGAGAAATTAATGAGTGTCCTTCGACGAGACAGTAACTCTGCGCGAA	909
Db	838	GTTCCCGCTGGCAAGAAATGCTGGTCCAGGCAATGCGCAAGAGAGGCGCTGTGCGCG	897
Oy	910	CGCTTCTTCGCTCGCATGAAGCTGTTCTTTCGCGGCGGCTGGGTTGTCCGAAGGATC	969
Db	898	AGCTTTTTCGCGACCTGMAATGCTGATGTATGTCGCGGTGCGGATGCGACAGACACA	957
Oy	970	TGGATCTGTTTGGACCGGGTGCCTGGAACAGACATGTGGTGAAGCGATTCGATGATGCG	1028
Db	958	TGGATGCGCTGAAGGAGCTTTCATAGGCAACGCTGGCCATTCGGGTTCTCATGGGTGC	1017
Oy	1030	GGTCTGGGCATGACGAGACGTCTCTTCTGCACTTTTACACCGGACCGCTGTGCATG	1088
Db	1018	GGCTCTCGGCTCGAGGAGACGGACCCCTTTCGCTGTCTGACACGAAACCCGACAGACAAG	1077

QY	1030	GGTCTGGGCGATGACGGAGACTGCTCTCTTGACATTTAACCAACGGACCGTCTCGAG	1089
Db	1018	GGCTCTGGCTCGACGGAGACGGCAACCTTCTCTGTTCTGCACCGAACCGACGACAG	1077

QY 1090 GCTGTTAATGAGGCTGCGGCGCTGCGAGGTCAAGCTGTTCCGATCGATGGG 1149
DB 1078 CCGGGCAATATCGGTATCCGGGCGAGGGCGTGAAGATGAACTCGGCTTCGATGGC 1137
QY 1150 AATTTGGAAGGGGCTTTCATGGTCCGACAGTCAATGAGCGGCTTACGCTGCTCTGAA 1209
DB 1138 CCGTACGAATTAAGGCTGAAAGGCCGACATCAAGCCCGGTTACGCGCTTAATGCGAA 1197
QY 1210 CAATATGCCCAAGCGTTCGACGAGGAGCTATTAAGTCCGAGTATGCTCAATTTG 1269
DB 1198 TTACGCGCGGCGACCTTCGACGAGGAGGTTCTATTCGAATCCGCAATCCGTAATTT 1257
QY 1270 GCAGATCTGCGCGATCCTCAAGAAAGTCTGATGTTGACCGGTCAATTTGCTGAAGCTTC 1329
DB 1258 GCGGTGCGGAGCATCCGCGCGCGCTTTATTTGATGGCCGCAATGGCGAAACTTC 1317
QY 1330 AAGCTGCTCAAGGGGATTTGTCAGGTTGGGCAATGGCGACGCGGCGGTTCTGAA 1389
DB 1318 AAGCTCAAGACCGGACCTGAGGCGGCTGCGGCGCTTTGCGCGCAATTAATGCAACATG 1377
QY 1390 GGGGCGCTTACGTCCTGAGCGTATGAGTGTGCTGCTCTGATGTAATGCTTGATTTG 1449
DB 1378 TTGCGCGGCGGTATCCGTATGCTGATGCTATACGCGGCAAAACCGGCGCAATGGGCGCT 1437
QY 1450 CTGCTGTTTCCGGGCTTCTCGAATGCGGCTGCTGCTGCGGCGTAAAGAAAGCGCTCG 1509
DB 1438 CTGCTGCTGCTATTCATACCGGCGCTACGTAAGCTGTTGCGGTTCCGAAACATCTCTCC 1497
QY 1510 GAGCGGAGGTCCTTGCCAGTGAAGCGGTTGCGGCGCTGTTGCTGACTGCTCAACGA 1569
DB 1498 GATGCGGAGATCATCCGCGACCCGAGTGTCAAGGCGCGCAATTCGCGAAGCTTTCGCGC 1557
QY 1570 CTCATTCGAGAACCACTGGCAATGCGATGCGATCATGAGGTAAGGCTCTCGATACG 1629
DB 1558 CATCAAAAGAGCGCGAGCGGTTCCGCAAGCGCGCTATGCGATTCCTGATGGAAGAC 1617
QY 1630 CCGCGCTCATGATTAAGAGGCGAGTCACTGACAGGCGCTGATCAACGCGCGCTGTT 1689
DB 1618 GCGCTGCGCTTCGAGAAAGGCGAGGTTACGGAACAAGGATGATCAACGCGCGCGTGT 1677
QY 1690 TTGCAATGGCGGTCGCGGAAAGTTGATGCGCTGTATCTGTG 1729
DB 1678 CTTCGTCATCGCAAGAGCTCGTAGATCATCTTACGCGCG 1717

RESULT 7
US-10-369-493-38427
/ Sequence 38427, Application US/10369493
/ Publication No. US20030233675A1
/ GENERAL INFORMATION:
/ APPLICANT: Cao, Yongwei
/ APPLICANT: Hinkle, Gregory J.
/ APPLICANT: Slater, Steven C.
/ APPLICANT: Goldman, Barry S.
/ APPLICANT: Chen, Xianfeng
/ TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
/ TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
/ FILE REFERENCE: 38-10(52052)B
/ CURRENT APPLICATION NUMBER: US/10/369,493
/ PRIOR FILING DATE: 2003-02-28
/ PRIOR APPLICATION NUMBER: US 60/360,039
/ NUMBER OF SEQ ID NOS: 47374
/ SEQ ID NO 38427
/ LENGTH: 1728
/ TYPE: DNA
/ ORGANISM: Agrobacterium tumefaciens
US-10-369-493-38427

Query Match 22.5%; Score 398.4; DB 15; Length 1728;
Best Local Similarity 52.8%; Pred. No. 2.2e-114;
Matches 908; Conservative 0; Mismatches 806; Indels 6; Gaps 2;

QY 13 GAGCGCTTCTTCCTTCCCGGGTCGAATTCCTGACGCTCTCGAGCATGGGCTAAGACC 72
DB 1 GATCGCTTGGCCCTATCCGAGAAATTAATGACGCTCTCTTCATGCTGTCTGTG 60
QY 73 CGTTCGAACAACTGCGTGTGCTGCGAGGCGGCAATGGGAAATGGGCTGTATACG 132
DB 61 GACCGGAGCGCAATGATGAGGCGGACGCTCAAGGTCGCGAGCATGGGCGCGGCTACG 120
QY 133 TACGCGAAATGTTCAACAACGTCGCGCATCGACAGAGCTTGTCTTCAAGCACTA 192
DB 121 TACGCGAGGCTTCACAAAGATACGCGGATGCGCAATGCTTCTTCATCATACCTT 180
QY 193 TCGCAGAGCGTCGCGCTTATCGTCTGGAATGACCTGGAACATTCACGTGCA 252
DB 181 TCGTGAGAGGCGCTGCTGCGGCTTCTGAAATTCATCGAGCATGGGTGATGATG 240
QY 253 TTTGGGCTATGATGCGGCGATTCCTTATGCGCGGTGTCTCTCTTATTAATGCTG 312
DB 241 CTTCGCGCGCGATGTCGATGCGCTGCGCTGCGCATACGCGCGCTATGCTC -- CAGC 297
QY 313 TCGCAAGATTTGGCGAAGCTGCGCTCATGCTGATGCTTCTGCAACCGGAGCTGCTT 372
DB 298 TCGCGCGATCTCACCAACTCGCGCAATCGCGGCGCAATACCCCGGAGATGTTT 357
QY 373 GCTGCGATGCAAGCACTTTCAGCGCGCAATGAGACATTCGCGGACGAGCTGCC 432
DB 358 GCGGAGATGCGACGCGCTTTCGCGCGGCTGCGGAGTTTTCAGACGAGTACCG 417
QY 433 GCAATCTTCACTGAGGCGAATTTGCGGCGCGCGC -- ACGTGATTTTGAACCTG 489
DB 418 CTGCTGCGCTCGCAACTGCGGAGATGCTTCATATGCTTCAATTCATTCGAAACCTG 477
QY 490 CTGAGCAGCTGCTGAGGATGAGGAGATTAATGCTTTCGCGCACTGCGCGCAATG 549
DB 478 CTTGAAACGAGCGGACCGAAGCGTTGATCGCTTTCGATGCGGTGCGGCGGATACC 537
QY 550 ATTCGCAAGTCTTGTTCACCTTGTGCTTACCAATGCTCAAGCGCGTCCGACTACT 609
DB 538 GTGCGCAATTCCTTTCACGTCAGGACGACTGTTGCCCAACCGGCTATCCAGACG 597
QY 610 CAGCGAATGCTGCGGCAATCAGAGATGCTTTCGCAACTTCCCGGTTTTTGTGAA 669
DB 598 CAACGATGTTGCTTCATCAGAAATGATGCGGATGCTTGTATGCTACTTTCGCGAG 657
QY 670 GAGCGCGGATGCTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 729
DB 658 GAGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 717
QY 730 ATCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 789
DB 718 TTCAACCTGCTCTATATATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 777
QY 790 CAAGGTTGCGGAGCGCTTTCGCACTGAGGAAATCTTTCGCACTGCGCTGCTGCTGCT 849
DB 778 GCGCAATGCGCGACGCTGCAATCTTCGGAATATGCGCACTGCTGCTGCTGCTGCTGCT 837
QY 850 GTGCGAAGAGCTGCGGAGAAATGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 909
DB 838 GTTCCGCGCTGCGCAGAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 897
QY 910 CCGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 969
DB 898 AGCTTTTTCGCGACCTGAAGATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 957
QY 970 TGGGATGCTTGGACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1029
DB 958 TGGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1017
QY 1030 GGTGCGGATGAGGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1089
DB 1018 GGTGCGGCTGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1077
QY 1090 GCTGTTAATTTGGCTGCGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1149

Db 1078 CCGGCAATTCGGTATCCGGCGCAAGGGCTGACGATGAAACTCGGCCCTTCGATGGC 1137
Qy 1150 AATTTGAAAGGGGCTTTCATGTGTCGACGTCATGAGCGGCTAATGCGCTGCTGAA 1209
Db 1138 CGGTACGAATTCAGGGCTGAAAGGCCGAAACATCACGCCGGTTACTGGCTAATGCGAA 1197
Qy 1210 CAAATGCGCAACCGCTTCGACGAGAAAGGCTATTACGTCCGGTGATGCTCAATTAATG 1269
Db 1198 TTGACGCGGACGCTTCGACGAGAAAGGCTTCATGCAATCGCGATACGTCAAATTT 1257
Qy 1270 GCAGATCTGCGCCGCTTCAGAAAGGCTGATGTTTGACCGTCAATTTGTAAGACTTC 1329
Db 1258 GCGGTGCGGACGATCCGCGCGCGGCTTTATTTGATGCGGCAATGCGGAAACTTC 1317
Qy 1330 AACCTGCTCAGAGGCTATTGTCAGCGTTGGGCAATGCGCAAGCGGCGGTTCTGAA 1389
Db 1318 AAGCTCAGACCGGCACTGGGTGGCGGCTTCGCGCGCAATTAGTCACATG 1377
Qy 1390 GGGGCTCTTAGGCTCGACGTAAGTGTGCTGCTCTGATGCTGAATGCTTGATG 1449
Db 1378 TTGCGCGGCTGATCCGTGATGCGGTATACGCGCGAAACCGGCGCAATGCGGCT 1437
Qy 1450 CTGCTGTTCCGCGCTTCTCGACTGCGCTGCTTGTGCGGCGCTAGGAAAGAGCGCTG 1509
Db 1438 CTGCTGCTGCTATTCATACCGGCGCTACGTGAGCTTGTGCGGTTGCGCAACTCTCC 1497
Qy 1510 GACGCGAGGCTTGGCACTGAGCCGTTGCGGCTGTTGCTGTAAGCTCAAAAGA 1569
Db 1498 GATGCGGAGATCATCGGCGACCGGAGTGTGACGGCGCAGATTGCGGAACTTTCGCG 1557
Qy 1570 CTGAATCGAAGCACTGGCAATGCAATGCAATGAGGAGGCTCTCGATAG 1629
Db 1558 CATCAAAACAGGCGAGCGGTTCCGCAAGCGGCTCATGCGCATTCGTGATGAAAGAC 1617
Qy 1630 CCGCCGCTCGATTGATTAAGGCGAGGCTCATGCAAGAGGCTCGATCAACAGCGCGTGT 1689
Db 1618 GCGCTGCGCTTGAGAAAGCGGATTAACGAGCAAGGATCGATCAACAGCGCGCTG 1677
Qy 1690 TTGCAATGCGGCTGCGGCAAAAGTTGATGCGCTGATCTG 1729
Db 1678 CTTCGATCGCAAGGAGCTGTAAGTCACTTACGCG 1717

RESULT 8
US-10-369-493-38903
; Sequence 38903, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 38903
; LENGTH: 1728
; TYPE: DNA
; ORGANISM: Agrobacterium tumefaciens
US-10-369-493-38903

Query Match 22.5%; Score 398.4; DB 15; Length 1728;
Best Local Similarity 52.8%; Pred. No. 2.2e-114;
Matches 908; Conservative 0; Mismatches 806; Indels 6; Gaps 2;
Qy 13 GAGGCGCTTCTCCCTCCGCGGTGCAATCTTGAGCGCTCGAGCATGGGCTAAGACC 72

Db 1 GATCCGCTGGCCCTTATCCGCAAAATTAATGAGGCTCTCTTACATGCTGCTTCG 60
Qy 73 GGTTCAGAACAAACCTGCGTGTCTGCCAGGCGGCAATAGGGATGCGCTGATACG 132
Db 61 GCACCGGAGCGCACATGAGATGGCGGACCGTCAGAGGTGCGCAAGCATGCGCCGGGTAGC 120
Qy 133 TAGCGGAATATGTTCCAAACGTCGCGGCATGCGACAGAGCTTGCTCTTACGSACTA 192
Db 121 TAGCGCAGGCTCTCGACAAAGATACGCGATGCGGCAAGTTCTTTCGATCATGACTT 180
Qy 193 TCGGAGAGCGTCCGCTGTTATGCTCTGAAATGACCTGGAACATCTTCACTGCGCA 252
Db 181 TCGGTGAGAGGCGCGGTGCTGATGCTTCTGAAAATTCATTCAGCATGCGTGTGATG 240
Qy 253 TTGGGGCTATGATGCGGGCAATCCCTATGCCCCGCTGTCTCTGTTATTCAGTCTG 312
Db 241 CTTCGCGCCAGCATGTGCGTATGCGCTCGCTGCTGCATCAAGCGCGCTATGCG--CAGC 297
Qy 313 TCGCAAGATTGAGGAAGCGGTGCAATGATGCTTCTGCAACCGGAGCTGCTT 372
Db 298 TCGGCGGATCTCAACAACTCGCGACATCGGCGGCAATCACCCGAGGATGTTT 357
Qy 373 GCTGCGGATCAGACACTTTCAGAGCGCAATTGAGACATTCCTGCGGAGAGCTGCCC 432
Db 358 GCGGAGATGCGAAGCCCTTTCGCGCGGCTGCGGAGGTTTTCAGACAGGATACACG 417
Qy 433 GCAATCTTCACTGAGGCGAATGGCCGCGGCGCG--ACGGTACTTTGACAGCTG 489
Db 418 CTGCTGCGCTCCGCAACCTGCGGAAAGATGTTCCAAATGTTTCAATTCGAAACCTTG 477
Qy 490 CTGAGAGCTGCTGAGGATGAGGAGATTAAGCTTTCGCGCACTGCGCCGATAG 549
Db 478 CTGGAACGAGCGGACCGAAGCGGTTGATGCTTTCATGAGGCTGCGGCGGATACC 537
Qy 550 ATTGCAAGTCTTGTTCATCTTCTGCTTACCAAACTGCTTAAGGCGGTGCGACTACT 609
Db 538 GTGCGCAATTCCTTTTCACTGCTCAGGACAGACTGTTGCGCAAGGCGTCAATCAGACG 597
Qy 610 CAGGAAATGCTTGGCGCAATACAGAGATGCTTCTGCAAACTTCCGGTTTTGTGAA 669
Db 598 CAACGATGTTGTCTTCAATCAGAAATGATGCGGATTTATGCTACTTTCGAG 657
Qy 670 GAGCGCGCGGTGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 729
Db 658 GAGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 717
Qy 730 ATCGGATGCTGTTGTAACAAGCGGCGGCTACTTACCTTTCAGGAGGTAACCAACGCC 789
Db 718 TTCACTGCTGCTCTATATATGCGGTACTTATATATGATGCGGCAAGCCAGCCG 777
Qy 790 CAAAGGTTGCGGAGAGCTTCCCACTTGAAGGAAATCTTCCACTGCTGACTACT 849
Db 778 GCGCAGATGCGGCGACACTGGAACAATCTCCGGGCAATATGCGGCACTGCTGTTTCAAC 837
Qy 850 GTGCGCAAGGCTGAGGAAATTAAGTGGTCCCTTGAAGGAGACAGTACCTCGCGGAA 909
Db 838 GTTCGCGTGGCCAGAAATGCTGTGTCAGGCAATGCGCAAGGAGAGGCGCTGTGCGC 897
Qy 910 CGCTTCTGCTGCTGCAAGAGCTGTTCTTTCGCGGCGGCTGCTGCTGCTGCTGCTGCTG 969
Db 898 AGCTTTTTCGCACTGGAAGATGCTGATATGCGGTGCGGCAATGCGGAGACACACA 957
Qy 970 TGGAGATGTTGACCGGCTGCTGAAACAGACTGTGTGAGCGCATTCGATGATGCGG 1029
Db 958 TGGAGATGCTGACGAGGCTTTCATGAGCAACGCTGCGGCAATGCTTCTGATGAGGCTC 1017
Qy 1030 GATCGGAGATGACGAGACTGCTCTCTGCTGCACTTTCACCGGAGCGGCTGCTGATG 1089
Db 1018 GCGCTGCGCTGACGAGAGGCACTTCTGCTGTTCTGCAACCAAGCGGAGACAG 1077
Qy 1090 GCTGTTTACATTTGGGCTGCCAGCGCTGCTGCGAGTCAAGCTGTTCCGGTCAATGGG 1149

Db	1078	CCGGGCAATATCGGTATCCGGGGCGAGGGGCTGACATGAATCGTGGCTTCGANTGGC	1137
Qy	1150	AAATTGGAAGAGGGCTTTCCATGGTCCGACAGTCATAGAAGGGCTACTGGCTGCTCTGAA	1209
Db	1138	CGGTACCAATTTAAGGGCTGAAAGGCCGGAACATACCCCGGTTACTGGGGTAAATGGCGAA	1197
Qy	1210	CAAAATCCCAAGCGTTTCGACGAGAAAGGCTATTACTGCTCCGGTGAATGCCATCAAAATTG	1269
Db	1198	TTGACGGCGGACACCTTTCACGAGAAAGGTTCTATCGAATGGGCAATACGTCAAAATT	1257
Qy	1270	GCAAGATCTGCGCATCTCTCAAAAAGTCTGAGTTTGAAGGTGAAATTGCTGAAGCTTC	1329
Db	1258	GCGGTGCGGACCATCTCGCCCGCGCTTTTATTTGAAATGGCGCATGCGGAAAACTTC	1317
Qy	1330	AAGCTGTCCTCAGAGGGATTTTGCAGCGTTGGGCGCATGCGGACGCGGGCGGTTCTGGAA	1389
Db	1318	AAGCTGCAGACCGGACACTGGGTGGCGCTGGGCGCTTTTGCGGCGCAATTAATGTAACATG	1377
Qy	1390	GGCGGCTCTTACGTCCTTGACGTAAGTGGTTGCTCTCTGTAATCGTGAATGCTTGAAATTG	1449
Db	1378	TTGCGCGGGGCTGATCCGTGATGCGGTATTAACGGGCGGAAACCGGGCGGACCTGGGCGCT	1437
Qy	1450	CTGCTGTTTCCGGCTCTTTCGACTGCGCGTGTCTGTGCGGGCTAGAGAAAAGGCGCTG	1509
Db	1438	CTGTCGTGTCATTCATCAACGGGCTACGTAAGCTTTGTCGGGTTTGGCAATCTCTCC	1497
Qy	1510	GAGCGCGAGGTGCTTTCGACATGAGCCGGTTTCGGGCTCTGGTTGCTGACTGCTCAAAAG	1569
Db	1498	GATGCGGAGATCATCCGGACCCGAGATGTCAGGGGCCCAATTTGTCGGAAGCTTTCCGGG	1557
Qy	1570	CTCAATCGAGAGCAACTGGCAATGCCAGTGCATCATGTGGGTAGAGGCTCTCTCGATAGC	1629
Db	1558	CATCAAAAACAGCGAGCGGTTTCCGCAAGCCGGCTATCGCATCTCTGTATGGAAGAC	1617
Qy	1630	CCGCGCGTCATTTAATAAGGCGAGGTCACTGAACAAAGGCTCGATCAACAGAGCGCTGTT	1689
Db	1618	GCGCTGCGCTTTCAGAAAGCGAGGTTTACGACAAAGGATCGATCAACAGCGCGCGCTG	1677
Qy	1690	TTGCAATGCGCGGTGCGGAAAGTTGATGCGCTGATCGTG	1729
Db	1678	CTTCTGTCATCGCAAGGAGCTCGTAGAGTCACTTTAAGGGG	1717
RESULT 9			
US-10-369-493-38007			
Sequence 38007, Application US/10369493			
Publication No. US20030233675A1			
GENERAL INFORMATION:			
APPLICANT: Cao, Yongwei			
APPLICANT: Hinkle, Gregory J.			
APPLICANT: Slater, Steven C.			
APPLICANT: Goldman, Barry S.			
APPLICANT: Chen, Xianfang			
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF			
FILE REFERENCE: 38-10(52052)B			
CURRENT APPLICATION NUMBER: US/10/369,493			
CURRENT FILING DATE: 2003-02-28			
PRIOR APPLICATION NUMBER: US 60/360,039			
PRIOR FILING DATE: 2002-02-21			
NUMBER OF SEQ ID NOS: 47374			
SEQ ID NO 38007			
LENGTH: 1660			
TYPE: DNA			
ORGANISM: Agrobacterium tumefaciens			
US-10-369-493-38007			

Db	1	GATCGCTTG6CCCCCTATCCGACGAAATGAAATGAGGCTCTCTTCACTGATGTCGTTG	60
Oy	73	CGTCCAGAAACAAACCTGACGTTGCTGCTCCAGGGCGGCAAAATGGGAATGGCGTGCATACG	132
Db	61	GCAACCGGAGGCGCACTGATGATGGCCGACCGCTCAGAGTGCAGGCACTATGGCCCGGCTAGC	120
Oy	133	TACGCGGAATATGTTCCACAAAGTCCGGCGCATCGACAGAGCTTGTTCTCTTACGAGCTA	192
Db	121	TACGCCAGAGGCTCTCGACCAAGATATCGGCGATATGGCCAGTTCTCTTCGATCATAGACTT	180
Oy	193	TCGGCAGAGCGTCCGCTGCTTATATGCTCTTGAAATGACTGGAACTTTCAAGCTGGCA	252
Db	181	TCGGTCCAGAGGCGCGCTGCTGCTGCTTTTGAAAATTCATCAGAGCATGCGTGAATGCTG	240
Oy	253	TTTGGGGCTATGATATGCGGGCATTTCCCTATTTGGCCGGGTGCTCTCGCTTTATCACTGCTG	312
Db	241	CTTTCGGGCCCAAGATATGCGGTATATGCTTCCGCTGCATCATCGCCGCGCTATGCTGCTG	297
Oy	313	TCGCAAGATTTGGGGAAGCTGCGTCACATGTAAGTCTTCTGCAACCGGAGACTGCTCTTT	372
Db	298	TCGGCCGATCTCACCAAACTGCGCGACATCCGCGGAGATATCAACCGGGGATGGTATTTT	357
Oy	373	GCTGCCATGACAGACCTTTTCCAGCGGCAATTGACACATTTCTGCCGACAGACGTGCC	432
Db	358	GCCAGGATCGAGACGCTTTTCCGCGGCGCTCGCGAGGTTTTCAGACGAGTACACCG	417
Oy	433	GCAATCTTCACTCGAGGCGCAATTGCGCGGCGGCGCTGCTGCTGCTGCTGCTGCTGCTG	489
Db	418	CTGCTCGGCTCTCGCAACCTGCGGGAAGATGTTTCCATATGTTTCAATTTTGAAACCTCTG	477
Oy	490	CTGAGACAGCTGCTGGGATTTGAGCGATATATGCTTTCGGCACTGGCCCCGATACG	549
Db	478	CTTGAAACCGAGCGCACCGAAGCGGTTGATCGTCTTCATGCGGTCCGGCCGGGATAC	537
Oy	550	ATTGCGCAAGTTCTTGTTTCACTTCTGCTGTACCAAACTGCTGTAGGCGGTGCCACTACT	609
Db	538	GTCGCGAAATTTCTTTTACAGTCAAGGACAGACTGCTGCTGCGCCAAAGCGGTATCCAGACG	597
Oy	610	CAGGAAATGCTCTGCGCAATCAGCAGATGCTTCTGCAAACTTTCCCGGTTTTTGTGAA	669
Db	598	CAAGCAATGTTGTCCTCAATCAGAAATGATTTGCCGATTGTTATGCTACTTTTCCGAG	657
Oy	670	GAGCGCGCGGTGCTGCTGAGCTGTTGCTGCTGGAACCAACCTTCCGCGCGACGCCAACAC	729
Db	658	GAGCGCGCGGTGCTGCTGATTTGAGTTGGCGCCATGGAACCAACCGCGCGGGAACAAAGTGT	717
Oy	730	ATCGGCATCTGTTGTATCAACGCGCGGCACTTACTTACAGACGGTAAACCAACCGCC	789
Db	718	TTCAACTCTGCTCTTATATATGGGGTACTTATATATGATGGCGGCAAGCCACGCGCG	777
Oy	790	CAGAGGTTGCGCGAGACGCTTCCGCACTTGAAGGAAATCTCTCCACTGCGTACTTACT	849
Db	778	GCGCAGATTCGACGACACTGGAACAATCTCCGGGACATATCGGCCCACTGCTATTTTCAAC	837
Oy	850	GTCGCGAAAGGCTGGGAGGAATATAGTGGTGGCCCTTGAAGGAGACAGTACCTCTGCGGAA	909
Db	838	GTTTCCCGCTGCGACGAAATGCTGCTCAGGCAATGCGCAAGACAGAGGCGCTGTGCGCG	897
Oy	910	CGCTTCTTCTGCTGCATGAAGCTGTTCTTCTTCCGCGCGGCTGAGTTTGTGCGAAGGATC	969
Db	898	AGCTTTTTCGCGACCTGGAAGATGCTGATGTATGCGGATCCGGAATGGCGAGACACA	957
Oy	970	TGGGATGTTTGAACCGGGTCCGCTGGAACAGACATGTGTGTGAGCGCAATTCAGATGATGGCG	1029
Db	958	TGGGATCGCTGACGAGGCTTTTCCATGGCAACGAGTCCGATGCGGTTCTCATATGGGTGCC	1017

	Query Match	21.0%	Score 372;	DB 15;	Length 1660;
	Beet Local Similarity	52.7%;	Pred. No. 4.5e-106;		
	Matches 877; Conservative	0;	Mismatches 780;	Indels 7;	Gaps 3.
QY	13 GAGGCGCTTCCCTTCGCCGGATTCGTAGCGCTTCGACATTGGGCTAAGACC	72			

Db 1018 GgCCTCGgCTCGAGGAGACgGgCACcCTTCTCGCTgTTCgCACCGAACCGCAGgACAG 1077

Oy 1090 GCTGTACATTTGGgGCTGCGAGCGGCTCGGCTGCGAGGTCAAGCTCGTTCGgTCGATgGg 1149

Db 1078 CCGGgCAATTCGGTATCCCGGCGGACAGgGCGTGCATGAATCTCGTGGCCCTTCATgGC 1137

QY 1150 AATTGAGAGGGGTTTCATGTCGACGTGATGAGCGGCTACGCGGCTCTGAA 1209
DB 1138 CGGTACGAATTAAGGCTGAAAGCCGAAATACGCGCGTATCTGCAATAGGCGAA 1197
QY 1210 CAATATGCCAAGGTTTGAAGAGGAGGCTATTAAGTCTCGGTGATGCAATATG 1269
DB 1198 TTGACGCGGAGCTTTCGAGAGGAGGAGGTTTCAATGAAATCGGCGATACCAATG 1257
QY 1270 GCGATCTCTCCGATCTCTGAGAAAGTCTGATGTTTGAAGGCTGCAATGCTGAACTTC 1329
DB 1258 GCGGTGCGGAGCATGTCGCGCGCGCTTTTATTCATGAGCCGCAATGCGGAAACTTC 1317
QY 1330 AAGCTGCTCAAGGAGTATTTGTCAGGTTGGGCGCATTTGCGAGCGGCGGCTTCTGAA 1389
DB 1318 AAGCTGAGACCGGACCTGAGGTCGCGGCTTTGCGCGGCAATTAATGATCAATG 1377
QY 1390 GCGGCTCTTACGTCCTGAGAGTATGCTGCTGCTGCTGCTGATGCTGATGCTTGAATG 1449
DB 1378 TTGCGCGGAGTATCGGTGATGCGGTATTAACGCGGCGAAACCGGCGCAATGCGGCT 1437
QY 1450 CTGCTGTTCCGCTCTTCTGACTGCGGCTTGTGCGGCGCTAGAGAAAGAGCGCTG 1509
DB 1438 CTGCTGCTGCTATTCATACCGGCGCTACGAGCTTGTGCGGCTTGCAGCAATCTCTCC 1497
QY 1510 GAGCGGAGGCTGCTGCAATGAGCGGCTTCCGCGCTTGTGCTGATGCTGCTCAACGA 1569
DB 1498 GATGCGAGATCATCCGCGACCCGAGTGTACAGGCGGAGATTTGCGGAACTTTCGCG 1557
QY 1570 CTCAATCGAAGCACTGCAATGCGATGCTGATCATGTCGAGGCTCTCTGATACG 1629
DB 1558 CATCAAAACAGGAGCGGTTCCGCAAGCGGCTGATGCGCATTCCTGATGAGAGAC 1617
QY 1630 CCGCGCTCATTTGATTAAGGCGAGGCTCACTGACAGGCGCTGAT 1673
DB 1618 GCGCTG-CGCTTCGAGAGGCGAGGTTACGACAAAGGATCGAT 1660

RESULT 10
US-10-369-493-40734
; Sequence 40734, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfang
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 40734
; LENGTH: 1866
; TYPE: DNA
; ORGANISM: Caulobacter crescentus
US-10-369-493-40734

Query Match 12.7%; Score 225.6; DB 15; Length 1866;
Best Local Similarity 48.1%; Pred. No. 5.9e-60;
Matches 775; Conservative 0; Mismatches 814; Indels 21; Gaps 4;

QY 118 TGCGCTGATCTACGACGAGAAATGTCACAACTCCGCGCATGCGACAGAGCTTG 177
DB 238 TGCGCGGCGCTACCTATGCGAGGCGCATGCGCGGTCAAGGCGCTGCGCCACTGCTG 297
QY 178 CTTCCTTACGAGATTCGAGAGAGCGCTGCGCTTATGCTCTCTGAAATGACCTGGA 237
DB 298 ATGATCAGAACTTACGAGCCCAAGACAGAGTATGATCTGTCGCGCAATTCGATGAG 357

QY 238 CATCTCAGCTGAGCTTGGGCGTATGATGCGGCAATTCCTATGCGCGGTCTCT 297
DB 358 CAGCCCTGATGAGCGCTGGCGCTTATACGACCGGCGCTGCGCGGCGGATAGCCCG 417
QY 298 GCTTATTCACGTGCTGTCGCAAGATTTGGCGAAGCTGCGTACATCTGAGTCTTCTGCA 357
DB 418 GCTCAGAGCTGATCTGACCGACACGCGCAAGTCAAGCATCTGCTTCAAAAGTGGCG 477
QY 358 CCGGAGCTGCTTTTGTGCTCCGATGACGACCTTTCAAGCGCGCAATTTGAGACCATTCG 417
DB 478 CCGCGCTCTCTTTCGCGCGAGCGAGCATGTTCCCGGCGCTGCGCACCTCTAAG 537
QY 418 CCGAGACGCTGCGCGCAATCTTCACTGAGGCGAAATTTGCGCGGCGCGCACGCTGAGT 477
DB 538 GCGCTGAGCCGAGCTCTGCTGATACCGCGATGCGGCGGCGAGGAGGCGATACCC 597
QY 478 TTGACAGCCTGTGAGACGCTGTGAGATTGAGCAGATTAATGCTTTGCGGCAACT 537
DB 598 TTGCGCGAGGTCCGCGGACCAAGCGGACCGCGCGCTGCGCGGCGGAGATGGGCTG 657
QY 538 GCGCGGATTCGATTCGCAAGTTCTTGTCACTTCTGCTTCACTCAAACTGCTTAAGCG 597
DB 658 GCGCGCGCGAGCTGCGCAAGTATCTTCACTCGGCTCAAGGAGTCTGCGCAAGGCC 717
QY 598 GTGCGCATCTACAGCGAATGCTCT--GCGCAATGAGAGATGCTTTCGCAACTTC 654
DB 718 GTGCGCGAGCCAGCGAGATGAGCGGCGCTATCCGCGCAAGAGGCGCTGCGCAC 777
QY 655 CCGGTTTTTGTGAGAGCGCGGCTGTGAGCTGAGATGTTGCGGTGGAACCACTTC 714
DB 778 GAGGTTCGAGCGACAGAGTTCGCCAGAGCTGGAATGATGCTGCGGACCATCTCG 837
QY 715 GCGGCGACCAACATCGGATGCTGTTTACACGCGCGCACGTACTACCTTGACGAC 774
DB 838 GCGGCGCAATCGGCTTCAACGCGGTGATCTGAGGCGCGGCGCACCTGCAATGACGAG 897
QY 775 GTTAAACCAACCGCCCAAGGTTTCGCGAGACGCTTTCGCAACTGAGCGAAATCTCTCC 834
DB 898 GCGAGCGCTGCGCGGATGTCGAGACCAACATCAAGAACCTTACGAGGTCTCCCG 957
QY 835 ACTGCGTACTCACTGTCGGAAGGCTGAGGAAATTAAGTGGTGCCTTGAAGCGAGAC 894
DB 958 ATGCTGTGCGGCTGCGCGCGGATCGCTTCTGATGCTGCGGAGGATGAGGAAGAC 1017
QY 895 AGTACCTGCGGAGACGCTTCTTCTGCTGCGATGAAGCTGTTCTTTCGCGCGGCTGCG 954
DB 1018 CTGTGCTGCGGCGCTGCTTCTTCAAGAACCTGCGTACATGAGGCTATGCGGCGCAC 1077
QY 955 TTGTCGCAAGGATCTGAGATCTTTGAGACCGGCTGCTGAACAGCATGCTGTGAGCGC 1014
DB 1078 CTAGCAATGATGCTTATGAGCGATTCAGGCGCTGCGCGGAGCGGCGCATGCGC 1137
QY 1015 ATTGCGATGAGCGGCTGCGCATGAGCGAGACTGCTCTTCTGCACTTTTACACACC 1074
DB 1138 ATTGCGTACCAAGATGACGCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 1194
QY 1075 GAGCCGCTGATGAGCTGTTACATTTGAGCTGCCAGCGCTGCTGCGAGGCTCAAGCTC 1134
DB 1195 TGATCACCGAGGCTGCGGCTGCTGCTGCGGCTGCGGCTGCGGCGGAGCTGAGAGCTG 1254
QY 1135 GTTCGCGTATGAGGAAATTTGGAAGGCGTTTCCATGCTCCGACAGCATGAGACGCTAC 1194
DB 1255 GCGCCAGCGGCTGAGAGTACGAGTCCGCTGTAAGGCGCGGAGCGGCGGCTAT 1314
QY 1195 TGCGTGTCTCTGAACAAATGCGCAAGCTTTCAGAGGAGGCTATTAAGCTCTCGGT 1254
DB 1315 CACAAGAACCGGAGGAGCGGCGGCGCTTTCAGAGGAGGCTTCTATTAAGCTGCGC 1274
QY 1255 GATGCGATCAAAATGAGATCTGCGCATCTCAAGAAAGTGTGATGTTTGAAGCTGCA 1314
DB 1375 GAGCGCGCGGCTGCTGATTCGAGACGACCCGCGCAAGGCGCTGCTTTCAGAGCGCGG 1434

QY 1315 ATTGCTGAAGACTTCAAGCTGCTCCAGAGGGATTTGTGACGCTTGAGCCATTGCCACG 1374
DB 1435 GTGACCGAGAGACTTCAAGCTCGACAGCGGTAAGCTGGGTCAGCGTCGAGGCGCC 1494
QY 1375 CGGCGGCTTGTGAAGCGCGCTTTCAGTCTTGAAGTGTGTTGCTGCTCTGATCGT 1434
DB 1495 GATCTGTGGCGCCGTCGACGCGCCCTTACATCAAGACGCGGTATCAACCGCCAGGAC 1551
QY 1435 GAATGCTTGAATGCTGCTGTTTCCGCGCTTCTGCACTGCGGCTTGTGCGGAGCTA 1494
DB 1552 -AAGGCTTTCATGCGGCGCAATGCTCTGCGCCCTCGCGGCGGCGGCTTATGAGC 1610
QY 1495 GGAAGAGCGCTGCGACCGCGAGGTCTTCCAGTGAAGCGGCTTGGGCTGCTTCT 1554
DB 1611 CGATCTGCGGCGCGAAGCGCGCTG-----GAGAGCTGATCGCATCTCAAG 1659
QY 1555 GACTGCTCAAGACTCAATCAAGAGCAAGCACTGCAATGCCATGCCATCATGTGGTA 1614
DB 1660 GACCGTCTGGCGCGCTTAAAGCGCGCGCGGCGCTTTCGCGGCGGCTGCGCGCTT 1719
QY 1615 GGGCTCTCGATAGCGCGCGCTGATTTGATTAAGGCGAGGTCACCTGACAGGCTCGATC 1674
DB 1720 ACATCTTCAACGACGCGCGCTGATCGACGCGCGGAGATCAACGACAGGCTTATC 1779
QY 1675 AACGACGCGCTGTTTTCGAATGCGCGCTGCGGAGAAAGTTGATGCGCTGTA 1724
DB 1780 AACGACGCTGCTGCTGAGCGCGCGGATGCTGAGGCGCTCTA 1829

RESULT 11
US-10-716-803-5
Sequence 5, Application US/10716803
Publication No. US20040229236A1

GENERAL INFORMATION:

APPLICANT: Blanc, Veronique
Blanche, Francis
Crouzet, Joel
Jacques, Nathalie
Lacroix, Patricia
Thibaut, Denis
Zagorec, Monique
Debusche, Laurent
De Crecy-Lagard, Valerie
TITLE OF INVENTION: Polypeptides Involved in The
Biosynthesis Of Streptogramins, Nucleotide Sequences
Coding For These Polypeptides And Their Use

NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:

ADDRESSEE: Finnegan, Henderson, Farbow, Garrett & Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.

COUNTRY: USA
ZIP: 20005-3315

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/716, 803
FILING DATE: 20-Nov-2003

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 09/635,359
FILING DATE: 09-AUG-2000
APPLICATION NUMBER: US 09/231,818
FILING DATE: 15-JAN-1999
APPLICATION NUMBER: US 08/403,852
FILING DATE: 10-MAY-1995
APPLICATION NUMBER: PCT/FR 93/00923
FILING DATE: 25-SEP-1993
APPLICATION NUMBER: FR 92/11441
FILING DATE: 25-SEP-1992

ATTORNEY/AGENT INFORMATION:
NAME: Arriago, Salvatore J.
REGISTRATION NUMBER: 46,063
REFERENCE/DOCKET NUMBER: 03806.0054-04000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4000
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1879 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: S.pristinaespiralis
FEATURE:
NAME/KEY: CDS
LOCATION: 110..1858
SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-10-716-803-5

Query Match 3.2%; Score 57.4; DB 18; Length 1879;
Best Local Similarity 56.7%; Pred. No. 5.5e-07;
Matches 106; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 1071 CACCGGACCGCTGTGATGGCTGTGATGTTGAGGCTGCGAGCGCTGCGAGGTCAA 1130
DB 1186 CACCGGACCGCTGTGATGGCTGTGATGTTGAGGCTGCGAGCGCTGCGAGGTCAA 1245
QY 1131 GCTGTTCCGCTGATGAGGAAATTTGAGGCGCTTTCATGTCGACGTCATGAGCG 1190
DB 1246 GCGCGTCCCGCGGAGAACCGGTGACGTCACCGCGGCTTACACGCTGCGCG 1305
QY 1191 CTACTGCGCTGCTCTGAAACAAATCCCAAGCGCTTTCAGAGAGAGCTATTACTGCTC 1250
DB 1306 CTACTGCGCGGCGCGGACGACACGCGCGCTTTCAGAGAGAGCTTTCACGCG 1365
QY 1251 CGGCTGAT 1257
DB 1366 CGGCGAT 1372

RESULT 12
US-10-338-110-119

Sequence 119, Application US/10338110

Publication No. US20040023254A1

GENERAL INFORMATION:

APPLICANT: Fuhrmann, Jeffrey J.
TITLE OF INVENTION: A Method To Assess Quorum Sensing Potential Of Microbial
TITLE OF INVENTION: Communities

FILE REFERENCE: HER-0056

CURRENT APPLICATION NUMBER: US/10/338,110

CURRENT FILING DATE: 2003-01-07

NUMBER OF SEQ ID NOS: 123

SOFTWARE: PatentIn version 3.2

SEQ ID NO 119

LENGTH: 536

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:
OTHER INFORMATION: Consensus Sequence

FEATURE:
NAME/KEY: misc_feature

LOCATION: (7)..(7)

OTHER INFORMATION: n is a, c, g, or t

FEATURE:
NAME/KEY: misc_feature

LOCATION: (21)..(21)

OTHER INFORMATION: n is a, c, g, or t


```

1 NAME/KEY: misc_feature
2 LOCATION: (33)..(33)
3 OTHER INFORMATION: n is a, c, g, or t
4 FEATURE:
5 NAME/KEY: misc_feature
6 LOCATION: (69)..(69)
7 OTHER INFORMATION: n is a, c, g, or t
8 FEATURE:
9 NAME/KEY: misc_feature
10 LOCATION: (87)..(87)
11 OTHER INFORMATION: n is a, c, g, or t
12 FEATURE:
13 NAME/KEY: misc_feature
14 LOCATION: (213)..(213)
15 OTHER INFORMATION: n is a, c, g, or t
16 FEATURE:
17 NAME/KEY: misc_feature
18 LOCATION: (226)..(226)
19 OTHER INFORMATION: n is a, c, g, or t
20 FEATURE:
21 NAME/KEY: misc_feature
22 LOCATION: (261)..(261)
23 OTHER INFORMATION: n is a, c, g, or t
24 FEATURE:
25 NAME/KEY: misc_feature
26 LOCATION: (273)..(273)
27 OTHER INFORMATION: n is a, c, g, or t
28 FEATURE:
29 NAME/KEY: misc_feature
30 LOCATION: (347)..(347)
31 OTHER INFORMATION: n is a, c, g, or t
32 FEATURE:
33 NAME/KEY: misc_feature
34 LOCATION: (357)..(357)
35 OTHER INFORMATION: n is a, c, g, or t
36 FEATURE:
37 NAME/KEY: misc_feature
38 LOCATION: (398)..(398)
39 OTHER INFORMATION: n is a, c, g, or t
40 FEATURE:
41 NAME/KEY: misc_feature
42 LOCATION: (419)..(419)
43 OTHER INFORMATION: n is a, c, g, or t
44 FEATURE:
45 NAME/KEY: misc_feature
46 LOCATION: (509)..(509)
47 OTHER INFORMATION: n is a, c, g, or t
48 FEATURE:
49 NAME/KEY: misc_feature
50 LOCATION: (517)..(517)
51 OTHER INFORMATION: n is a, c, g, or t
52
53 OS-10-338-110-119

```

Query Match	2.6%	Score 46.8	DB 16	Length 536
Best Local Similarity	11.7%	Pred. No. 0.00077		
Matches	54	Conservative 220	Mismatches 187	Indels 0
			Gaps	0
Oy	382	GCAGACCTTTTCACGCGGCAATTGAGACCATTTTCGCCGAGACGCTGCCCGCAATCTTC	441	
Db	73	SVRSYBYCWMYSKNMNVVGDMMBSBKMGVARYKBGAYCATWCGAAMVYVMSASHRSSKC	132	
Oy	442	ACTGAGGCGAATTGGCCGGCGGGCGGACCGAGAGTTTGACAGCCTGTGTGAGCAGCGCT	501	
Db	133	TACTVSBTSGSHKHKMGVSVYVGDHATSKRYGHTGTGCGSGSTGHTDCCBACSACV	192	
Oy	502	GGTGGGATTGAGGCAATTAATGCTTTTGGCGGCACTGGCCCCGATACGATTGCCAAGTTC	561	
Db	193	BDSGCBMAVCTGCTBRARKVNVSTVTTYRSBMVNYNSBKCBVSRVRYVMVYRCCSMVSVAT	252	
Oy	562	TTGTTCACCTTGTGCTTACCAAACTGCTTAAGCGGAGTCCGACATCTACTAGGAATGTC	621	
Db	253	CCVMKSRNTTGGGAHTKMCNGCMTWCRSCRCBRCRHSCSSAMVXGGSIRGMVWMRY	312	

Qy	682	CTGGTGAACGTGTTGGCCGTGAACACACACCTTCGGCGGAGCACAACATCGGCATCGTG	7411
Db	313	TSASSVMMBCSTVAARMSCGSDSYVWSBSGGNGCYVRBVSNTSRTGGSVRYKYSW	3722
Qy	622	TGGCCCAATCACCAAGATCGTTCTGCAAACTTCCCGATTTTGTGTGAAGAACCCCGGATG	6611
Db	313	TSASSVMMBCSTVAARMSCGSDSYVWSBSGGNGCYVRBVSNTSRTGGSVRYKYSW	3722
Qy	742	TTGTACAAAGCGGGCAGCATCTTACGACGCTAAACCAACCGCCCAAGGTTGGCC	8010
Db	433	YBNHSSVMMHVRBRBRBDNRVTSTSGCSTSYYGAAACRVADACRMDSCVYRMVSIBY	4929
Qy	802	GAGACGCTTCGCACTTGAGCGCAATCTCTCCCACTGGGTA	842
Db	493	KYWSCKTYRSSMHIYCVNHRHBGVBMNVKRYGVGVDRYKY	533

RESULT 13
 US-10-369-493-42307
 Sequence 42307. Application US/10369493
 Publication No. US2005023675A1
 GENERAL INFORMATION:
 APPLICANT: Cao, Yongwei
 APPLICANT: Hinkle, Gregory J.
 APPLICANT: Slater, Steven C.
 APPLICANT: Goldman, Barry S.
 APPLICANT: Chen, Xianfeng
 TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 FILE REFERENCE: 38-10(52052)B
 CURRENT APPLICATION NUMBER: US/10/369,493
 CURRENT FILING DATE: 2003-02-28
 PRIOR APPLICATION NUMBER: US 60/360,039
 PRIOR FILING DATE: 2002-02-21
 NUMBER OF SEQ. ID NOS: 47374
 SEQ. ID NO 42307
 LENGTH: 1566
 TYPE: DNA
 ORGANISM: Halobacterium sp. NRC-1
 US-10-369-493-42307

Query Match	2.6%	Score 46.8	DB 15	Length 1566
Best Local Similarity	49.2%	Pred. No. 0.0011		
Matches 123	Conservative 0	Mismatch 127	Indels 0	Gaps 0

Qy	517	GATATATCCTTTGGCGCACTGGCCCGGATACGATTGGCAAGTTCTTGTTCACTTTGGCC	576
Db	439	GACGACACGGTTGGCGCCCGCGACGACGACCTCGCCGTCAGACGGTACACCTCCGGC	498
Qy	577	TCTTACCAACTGCTTAAGCGCGTCCGACTACTACGGAATGCTTCTGCGCCATTCAGAC	636
Db	499	ACCAACCGGACGCCCAAGGGGTCACTACACACCAACACTTCGCGTCCAGCGCCGAC	558
Qy	637	ATGCTTTGCAAACTTTCCGGTTTTTGATGAAGCCGCGGTGCTGGTGAAGTGGTTG	696
Db	559	CAGTCCATGACATCGTGCCTCGACGGCTCCAAACCGCGAGCCGCAAGATCGGGGTG	618
Qy	697	CCGTGAACCAACACTTCGGCGCGACGACCAACAATCGGCATGTGTTGTACACGCGCGC	756
Db	619	CCGTTGTTCCACATCTACGCGATGACCGTCGGCATGAACGCCGCGCTGTTGCGCGGCGC	678
Qy	757	ACGFACTACC	766
Db	679	ACGTTCTACC	688

RESULT 14
US-10-282-122A-30430
; Sequence 30430, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu


```

Db      241  |||||CTTCAGCTGACATTTGGGGCTATGATGCGGGCATTCCTATTTGCGGGGTCTCCGCT 300
Qy      301  |||||TATTCATGCTGTCGCAAGATTTTGGCAACCTGCTGCATCATGCTAGATCTTTTGGCAACG 360
Db      301  |||||TATTCATGCTGTCGCAAGATTTTGGCAACCTGCTGCATCATGCTAGATCTTTTGGCAACG 360
Qy      361  |||||GACATGCTCTTCTGCTGCGATGAGCACCCTTTCAGCGCGCAATTGAGACCATTTGCGC 420
Db      361  |||||GACATGCTCTTCTGCTGCGATGAGCACCCTTTCAGCGCGCAATTGAGACCATTTGCGC 420
Qy      421  |||||GACGAGTGTCCCGCAATCTTCACTCGAGCGAAATTGCGCGCGCGCAAGTGTGTTT 480
Db      421  |||||GACGAGTGTCCCGCAATCTTCACTCGAGCGAAATTGCGCGCGCGCAAGTGTGTTT 480
Qy      481  |||||GACAGCTGTGAGAGCAGCTGTGGGATTTGAGGAGATTAATGCTTTTGGGCAACTGAGC 540
Db      481  |||||GACAGCTGTGAGAGCAGCTGTGGGATTTGAGGAGATTAATGCTTTTGGGCAACTGAGC 540
Qy      541  |||||CCGATTCAGATTCGCAAGTCTTGTTCATCTTGTGCTCTACCAAACTGCTTAAGCGGTG 600
Db      541  |||||CCGATTCAGATTCGCAAGTCTTGTTCATCTTGTGCTCTACCAAACTGCTTAAGCGGTG 600
Qy      601  |||||CCGATTCAGAGGATGCTCTGCGCAATCAGCAGATGCTTCTGCAAACTTTCGCGTT 660
Db      601  |||||CCGATTCAGAGGATGCTCTGCGCAATCAGCAGATGCTTCTGCAAACTTTCGCGTT 660
Qy      661  |||||TTTGGTGAAGCGCGCGGTGCTGTGGAATGCTGTGCGGGAACCACTTTCGCGCGC 720
Db      661  |||||TTTGGTGAAGCGCGCGGTGCTGTGGAATGCTGTGCGGGAACCACTTTCGCGCGC 720
Qy      721  |||||AGCCACAATCGGATCGATGTTGTATCAACGCGGCGCATCTTGAAGAGGTAA 780
Db      721  |||||AGCCACAATCGGATCGATGTTGTATCAACGCGGCGCATCTTGAAGAGGTAA 780
Qy      781  |||||CCAAACCGCCCAAGGCTTCCCGAGACGCTTTCGCACTTGAAGCAATCTTCCCATGCG 840
Db      781  |||||CCAAACCGCCCAAGGCTTCCCGAGACGCTTTCGCACTTGAAGCAATCTTCCCATGCG 840
Qy      841  |||||TACCTCACTGTGCGAAGGCTGGGAGATTTGTGGGTGCGCTTGAAGCAGTACC 900
Db      841  |||||TACCTCACTGTGCGAAGGCTGGGAGATTTGTGGGTGCGCTTGAAGCAGTACC 900
Qy      901  |||||CTGCGCGAAGCTTCTTCTGCTGCGATGAGCTGTTCTTCTTCTGCGCGCGCTGGGTTGCG 960
Db      901  |||||CTGCGCGAAGCTTCTTCTGCTGCGATGAGCTGTTCTTCTTCTGCGCGCGCTGGGTTGCG 960
Qy      961  |||||CAAGGATCTGGGATCGTTTGAACCGGGTCTGTAACAGCATGCTGTGAGCGCATTCGCG 1020
Db      961  |||||CAAGGATCTGGGATCGTTTGAACCGGGTCTGTAACAGCATGCTGTGAGCGCATTCGCG 1020
Qy      1021  |||||ATGATGCGGGGTCTGGGATGAGGAGATGCTCTTCTGCACTTTTACACCGGACG 1080
Db      1021  |||||ATGATGCGGGGTCTGGGATGAGGAGATGCTCTTCTGCACTTTTACACCGGACG 1080
Qy      1081  |||||CTGTGATGCTGTTTACATTTGGGCTGCGCAGCGCTGCTGCAAGCTTCTGCG 1140
Db      1081  |||||CTGTGATGCTGTTTACATTTGGGCTGCGCAGCGCTGCTGCAAGCTTCTGCG 1140
Qy      1141  |||||GTGATGAGGAAATTTGAAAGGCGCTTTCATGCTCCGACGCTCATGAGCGCTTACGCGCT 1200
Db      1141  |||||GTGATGAGGAAATTTGAAAGGCGCTTTCATGCTCCGACGCTCATGAGCGCTTACGCGCT 1200
Qy      1201  |||||GCTCTGAAACAAATGCGCAAGGCTTTCAGAGGAGGCTTATTCCTCCGATGCG 1260
Db      1201  |||||GCTCTGAAACAAATGCGCAAGGCTTTCAGAGGAGGCTTATTCCTCCGATGCG 1260
Qy      1261  |||||ATCAAAATGCGAGATCTTCCGATCTCAGAAAGCTTGAATGTTGAAGCTGCAATTCCT 1320
Db      1261  |||||ATCAAAATGCGAGATCTTCCGATCTCAGAAAGCTTGAATGTTGAAGCTGCAATTCCT 1320
Qy      1321  |||||GAAGACTTCAAGCTGCTCAGAGGATTTTGTGAGCGTTGGGCACTTTCGACGCGGGCG 1380

```

```

Db      1321  |||||GAAGACTTCAAGCTGCTCAGAGGATTTTGTACGCTTGGGCACTTGGCAGCGGGCG 1380
Qy      1381  |||||GTTCTGAAGGCGGCTCTTACGCTCTGAGACGATGCTGCTGCTGATCGTGAATGC 1440
Db      1381  |||||GTTCTGAAGGCGGCTCTTACGCTCTGAGACGATGCTGCTGCTGATCGTGAATGC 1440
Qy      1441  |||||CTTGATGCTGCTGTTTCCGCGCTTCTGACCTGCGGCTTGTGCGGGCTAGGAAAA 1500
Db      1441  |||||CTTGATGCTGCTGTTTCCGCGCTTCTGACCTGCGGCTTGTGCGGGCTAGGAAAA 1500
Qy      1501  |||||GAGGCTCGGAGCGCGGCTTGTGCAATGAGCGCGCTTCCGCGCTGTTGCTGACTGG 1560
Db      1501  |||||GAGGCTCGGAGCGCGGCTTGTGCAATGAGCGCGCTTCCGCGCTGTTGCTGACTGG 1560
Qy      1561  |||||CTCAACGACTCAATGAGAAAGCACTGCAATGCGCATGCTGATCATGTGAGTGGAGCTC 1620
Db      1561  |||||CTCAACGACTCAATGAGAAAGCACTGCAATGCGCATGCTGATCATGTGAGTGGAGCTC 1620
Qy      1621  |||||CTCGATACGCGCGCGCTGATTAAGGCGAGCTCACTGACAAAGGCTCGATCAACAG 1680
Db      1621  |||||CTCGATACGCGCGCGCTGATTAAGGCGAGCTCACTGACAAAGGCTCGATCAACAG 1680
Qy      1681  |||||CGGCTGTTTGAATGCGGCTCGGGAAGTTGATGCGCTGATGCTGATGATGAAATCAA 1740
Db      1681  |||||CGGCTGTTTGAATGCGGCTCGGGAAGTTGATGCGCTGATGCTGATGATGAAATCAA 1740
Qy      1741  |||||TCCATGCTGCTGACGAGGCACTGATGA 1770
Db      1741  |||||TCCATGCTGCTGACGAGGCACTGATGA 1770

```

```

RESULT 2
US-08-976-063E-1
; Sequence 1, Application US/08976063E
; Patent No. 6524831
; GENERAL INFORMATION:
; APPLICANT: Steinbuehl, Alexander
; APPLICANT: Pfelefer, Horst
; APPLICANT: Rabenhorst, Jurgen
; TITLE OF INVENTION: SYNTHETIC ENZYMES FOR THE PRODUCTION OF CONIFERYL
; TITLE OF INVENTION: ALCOHOL, CONIFERYLALDEHYDE, FERULIC ACID, VANILLIN AND
; TITLE OF INVENTION: VANILLIC ACID AND THEIR USE
; FILE REFERENCE: Bayer-9998-CAO
; CURRENT APPLICATION NUMBER: US/08/976,063E
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: 196 49 655.1 GERMANY
; PRIOR FILING DATE: 1996-11-29
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Patentlin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 32679
; TYPE: DNA
; ORGANISM: Pseudomonas sp.
; FEATURES:
; NAME/KEY: CDS
; LOCATION: (3146)..(3997)
; OTHER INFORMATION: gene = "ORF1"
US-08-976-063E-1

```

```

Query Match      100.0%; Score 1770; DB 4; Length 32679;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1770; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1  |||||ATGCGTTCCTGAGGCGCTTCTTCCCTTCCGCGGTCGATTTCTTGAAGCTTCGAGCAT 60
Db      21288  |||||ATGCGTTCCTGAGGCGCTTCTTCCCTTCCGCGGTCGATTTCTTGAAGCTTCGAGCAT 21347
Qy      61  |||||TGGGCTTAAGACCGCTGCAAGAACTGCTGCTGCAAGGCGGCAATGAGGAATGG 120
Db      21348  |||||TGGGCTTAAGACCGCTGCAAGAACTGCTGCTGCAAGGCGGCAATGAGGAATGG 21407
Qy      121  |||||CGTGTATCAGCTTACGCGGAAATGTTCCACAAGTCCGCGCATTCGACAGACTTGCTT 180

```

```
Db 21408 CGTGTATCAGCTACGGGAAATGTTCCACAACGCTCCGCGCATCGACAGAGCTTCTT 21467
Qy 181 CTTTACGCACTATCGGAGAGCGTCCGCTGTTATGCTCTGTGAAATGACCTGGAACAT 240
Db 21468 CTTTACGCACTATCGGAGAGCGTCCGCTGTTATGCTCTGTGAAATGACCTGGAACAT 21527
Qy 241 CTTTACGCACTATCGGAGAGCGTCCGCTGTTATGCTCTGTGAAATGACCTGGAACAT 300
Db 21528 CTTTACGCACTATCGGAGAGCGTCCGCTGTTATGCTCTGTGAAATGACCTGGAACAT 21587
Qy 301 TATTTACCTGCTGTCGCAAGATTTGGCGAAGCTGCGTCACTATGAGTCTTCTGCAACCG 360
Db 21588 TATTTACCTGCTGTCGCAAGATTTGGCGAAGCTGCGTCACTATGAGTCTTCTGCAACCG 21647
Qy 361 GAACTGCTCTTTGCTGCTGCAATGCAACCTTTCCAGCGCGCAATTGAGACCATTTCTGCG 420
Db 21648 GAACTGCTCTTTGCTGCTGCAATGCAACCTTTCCAGCGCGCAATTGAGACCATTTCTGCG 21707
Qy 421 GAGGAGCTGCGCGCAATCTTCACTCGAGGCGAATTGGCGCGGCGCGCAAGGTGATTTT 480
Db 21708 GAGGAGCTGCGCGCAATCTTCACTCGAGGCGAATTGGCGCGGCGCGCAAGGTGATTTT 21767
Qy 481 GAGGAGCTGCGCGCAATCTTCACTCGAGGCGAATTGGCGCGGCGCGCAAGGTGATTTT 540
Db 21768 GAGGAGCTGCGCGCAATCTTCACTCGAGGCGAATTGGCGCGGCGCGCAAGGTGATTTT 21827
Qy 541 CCGGATTCGATTTGCGCAAGTCTTGTTCATCTTGGCTTCAACAACTGCTTAAGGCGGTG 600
Db 21828 CCGGATTCGATTTGCGCAAGTCTTGTTCATCTTGGCTTCAACAACTGCTTAAGGCGGTG 21887
Qy 601 CCGGATTCGATTTGCGCAAGTCTTGTTCATCTTGGCTTCAACAACTGCTTAAGGCGGTG 660
Db 21888 CCGGATTCGATTTGCGCAAGTCTTGTTCATCTTGGCTTCAACAACTGCTTAAGGCGGTG 21947
Qy 661 TTTGGTGAAGAGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
Db 21948 TTTGGTGAAGAGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 22007
Qy 721 AGCCACAACATCGGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
Db 22008 AGCCACAACATCGGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 22067
Qy 781 CCAACGCGCCAAAGGCTTCCGCGAAGCTTCCGCACTTGAAGCAAACTCTCCCACTGCG 840
Db 22068 CCAACGCGCCAAAGGCTTCCGCGAAGCTTCCGCACTTGAAGCAAACTCTCCCACTGCG 22127
Qy 841 TACTCACTGTCGGAAGGCTGCGGAGAAATTAAGTGGTCCCTTGAAGCAAGTAAACC 900
Db 22128 TACTCACTGTCGGAAGGCTGCGGAGAAATTAAGTGGTCCCTTGAAGCAAGTAAACC 22187
Qy 901 CTGGCGCAAGCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
Db 22188 CTGGCGCAAGCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 22247
Qy 961 CAAGGAGATCGGGAATGCTTTGGAACGCGGCTGCTGAAGCACTGCTGCTGCTGCTGCTGCT 1020
Db 22248 CAAGGAGATCGGGAATGCTTTGGAACGCGGCTGCTGAAGCACTGCTGCTGCTGCTGCTGCT 22307
Qy 1021 ATGATGCGGAGTCTGCGCATGCGAGACTGCTCTTCTGCACTTTTACCAACCGGACCG 1080
Db 22308 ATGATGCGGAGTCTGCGCATGCGAGACTGCTCTTCTGCACTTTTACCAACCGGACCG 22367
Qy 1081 CTGCTGATGCTGCTTCACTTTGGGCTGCGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
Db 22368 CTGCTGATGCTGCTTCACTTTGGGCTGCGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 22427
Qy 1141 GTGATGAGAAATGGAAGGCGTTTCCATGCTGCGACGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
Db 22428 GTGATGAGAAATGGAAGGCGTTTCCATGCTGCGACGCTGCTGCTGCTGCTGCTGCTGCTGCT 22487
Qy 1201 GCTCTGTAACAAATGCGCAAGCTTTCAGACGAGAGGCTATTAATGCTGCTGCTGCTGCTGCT 1260
Db 22488 GCTCTGTAACAAATGCGCAAGCTTTCAGACGAGAGGCTATTAATGCTGCTGCTGCTGCTGCT 22547
```

```
Qy 1261 ATCAAAATGCGAGATTCCTGCGATTCCTCAGAAAGTCTGATGTTTGAAGGCTCGAATTGCT 1320
Db 22548 ATCAAAATGCGAGATTCCTGCGATTCCTCAGAAAGTCTGATGTTTGAAGGCTCGAATTGCT 22607
Qy 1321 GAAGACTTCAAGCTGCTCTGAGGAGGATTTGTGACGCTTGGGACATTCGACGCGGCG 1380
Db 22608 GAAGACTTCAAGCTGCTCTGAGGAGGATTTGTGACGCTTGGGACATTCGACGCGGCG 22667
Qy 1381 GTTCTGGAAGCGGCTCTTACGCTGCTGAGACGTAAGTGTGCTGCTGCTGCTGCTGCTGCT 1440
Db 22668 GTTCTGGAAGCGGCTCTTACGCTGCTGAGACGTAAGTGTGCTGCTGCTGCTGCTGCTGCT 22727
Qy 1441 CTTGGAATGCTGCTGCTTCCGCTCTTCTGCACTGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
Db 22728 CTTGGAATGCTGCTGCTTCCGCTCTTCTGCACTGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 22787
Qy 1501 GAGGCTGCGAGCGCGGCTGCTTGGCAAGGAGGCTGCTGCAAGGCTGCTGCTGCTGCTGCTGCT 1560
Db 22788 GAGGCTGCGAGCGCGGCTGCTTGGCAAGGAGGCTGCTGCAAGGCTGCTGCTGCTGCTGCTGCT 22847
Qy 1561 CTCGAAGCACTCAATCGAAGAGCACTGCGCAATGCTGCAATCATGCTGCTGCTGCTGCTGCT 1620
Db 22848 CTCGAAGCACTCAATCGAAGAGCACTGCGCAATGCTGCAATCATGCTGCTGCTGCTGCTGCT 22907
Qy 1621 CTTGATTCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680
Db 22908 CTTGATTCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 22967
Qy 1681 CGGCTGCTTGTGCAATGCGGCTGCGCGCAAGTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
Db 22968 CGGCTGCTTGTGCAATGCGGCTGCGCGCAAGTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 23027
Qy 1741 TCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1770
Db 23028 TCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 23057

RESULT 3
US-09-328-352-2393
; Sequence 2393, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNIT FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 2393
; LENGTH: 1902
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-2393

Query Match 25.2%; Score 445.4; DB 4; Length 1902;
Best Local Similarity 55.2%; Pred. No. 1.5e-118;
Matches 932; Conservative 0; Mismatches 746; Indels 9; Gaps 3;

Qy 46 GAGGCTGAGAGATGCGCTGCTATCAGCTACGCGGAAATGTTCCAAACGTCGCTGCTGCGGCGCATC 165
Db 166 GATGCTTTAATTCATTTTGGCAAAACCAACCCGACATATTTTGGACGAAAAAGCAAT 225
Qy 106 GCAAAATGGAATGCGCTGCTATCAGCTACGCGGAAATGTTCCAAACGTCGCTGCGGCGCATC 165
Db 226 GCTCAAGGCAATGCGCTGCTATCAGCTACGCGGAAATGTTTCCAAACGTCGCTGCGGCGCATC 285
Qy 166 GCACAGAGCTTCTGCTTACGAGCTATGCGAGAGAGTCCGCTGCTATGCTCTGCA 225
Db 286 GCTCAAGCTTCTGCTTACGAGCTATGCGAGAGAGTCCGCTGCTATGCTCTGCA 345
Qy 226 AATGACCTGGAACATTTTCAGCTGCAATTTGGGCTATGATGCGGCGATTCCTATTCG 285
```

Db 346 AATGATCTCGAATTTAACAATGCTATGCTGCGCAATGCTGCGAGCGCTTTCTCG 405
 Qy 286 CCGGTCTCTCTCTATTAATCACTGCTGCGAAGATTTGGCGAAGCGCTCACTGTA 345
 Db 406 GCTATTTCCCTCTACTCTCTGATTTCTCAAGACTTTGGCAACTGCAAACTGCTTT 465
 Qy 346 GGTCTTCTGCAACCGGAGCTGTCTTGTCTGCGAATGCAAGCACTTTCCAGCGCAAT 405
 Db 466 GAAGTGTCAACCTGGTATGTTATGCGAGGATGGACAGACTTTTGGCAAGGATTT 525
 Qy 406 GAAGCAATTTGCGCGAAGCAAGCGCGCAATCTTCACTGAGCGCAATTTGGCGGCGG 465
 Db 526 CAGGCAATGATTAAGCTGATTAATGAAGTGAACCAATGAAGAAATAGGCGCGATCAG 585
 Qy 466 CGAGCGGTGATTTTGAAGCTGCTGAGAGCAAGCTGGTGGATTTGAGCGAATTAATGCC 525
 Db 586 ATTTGCAAGCTTTTCAATGCTGTATGATTAACCAAGTTTCAAAATGTTCAAGGTTT-- 642
 Qy 526 TTTCGCGCACTGCGCCGATTAAGATTTGCAAGTTCTTGTCACTTTGCTCTACCAAA 585
 Db 643 TATCAAAACCTTATGATAAACAAGATTGCAAAATTTCTGTTTCACTAGGTTCACTAAA 702
 Qy 586 CTGCTTAAGCGGTGCGGACTACTCAGCGAATGCTGCGCCCAATCAGCAATGCTTCTG 645
 Db 703 TTACTTAAGCTGTACCGACACACATTTAATGTGTGTATATCAGCAAAATGTTAATG 762
 Qy 646 CAATCTTCCGCTTTTGTGTAAGAGCGCGGTGCTGAGATTTGCTGCGGTGCGGTGAAC 705
 Db 763 CAGACTTTCCCTAGATTGTAAGAAACACCGCTGTCTTACTCGACTGCGCTTGTGACAC 822
 Qy 706 CACACCTTCCGCGCGACCAACCAATCGGATGCTGTGTTGTAACAAGCGCGCACTACTAC 765
 Db 823 CACACATTTTGGCGGAGTCACAATGTCGATCGCACTGTAATAACGCGGTACAGATTAC 882
 Qy 766 CTGGAAGCGGTAAACCAACCGCCCAAGGTTGCGCGAAGCGCTTGCACTTGAAGGAA 825
 Db 883 ATGATGATGCAAAACCCGTTGCAAGAAATTTGACAAACTTATCTTAATCTCAAGAA 942
 Qy 826 ATCTCTCCCACTGTAACCTCACTGTCGCAAAAGGCTGGAGGAATTAAGGCGTTCCTT 885
 Db 943 ATTTCTCCAACTGTTATTTAAATGTGCCAAAGTTGGGAAGAACTCACCGAAGCTTA 1002
 Qy 886 GAGCGAAGACGTACCTGCGCGAAGCGTTTCTGCTGCAATGAAGCTTCTTCTGCG 945
 Db 1003 GAAAGATGAAGATTAAGAGCGCTTTTGGCCAAAGTTAAATTTATCTTTGCC 1062
 Qy 946 GCGGCTGGGTGTCGCAAGGATCTGGATCTGTTGACCGGCTGCTGAACAGCACTGT 1005
 Db 1063 GGTGCTGCACTTTCAAGAGCGGCTGGAACAGACTGATTAATTTGCTCAGCAACTTGC 1122
 Qy 1006 GGTGAGCGCAATTCGATGATGCGGCTGCGGATGACGAGACTGCTCTTCTGCACT 1065
 Db 1123 GGAAGAAAATTCGCACTTAATGAGCGGATTTGGCAATGACGAACTGCTCTTGTGCT 1182
 Qy 1066 TTATACACCGGACCGCTGCTGATGCTGATTAATTTGCGCTGCAAGCGCTGCTGCGAG 1125
 Db 1183 TTATACAACTGCGCAAGCGGTAAATGGCTGCTTATTGTTAACCCTGCTCGGGATGGAA 1242
 Qy 1126 GTCAAGCTGCTTCCGCTGATGATGGAATTTGAAGGCGGCTTCAATGCTCGCACTGATG 1185
 Db 1243 ATTAACTGATTTCCATGCTGTGCAAACTGAGTTTTCGCTGTCGCAAAACGCTCATG 1302
 Qy 1186 AGCGGCTACTGGCG--TGCTCTGAAACAATATGCCAAGCGTTTCGACGAGAGGCTAT 1242
 Db 1303 AAAGGCTATTGGCGCTTAAAGGCGACCAACAAGTACTATTTTGTATGATGAAGGCTTT 1362
 Qy 1243 TACTGCTCCGCTGATGCAATTAATGGCAATCTGCGCATCTCGAAGAGCTGATG 1302
 Db 1363 TTCCATTAAGAGGCGATGCGCTTGTGATGCTGATGCAATGATCAACCAAGGCGTAAATG 1422
 Qy 1303 TTGAGCGGTGATGCTGAAGCTTAAGCTGCTCAGGGGATTTTGTCAAGGCTTGGG 1362

Db 1423 TACAGCGAAGATTTGCCAAGACTTTAACTCAATACAGGCACTTTGTGAATGTCGC 1482
 Qy 1363 CCATTCGCGACGCGGCGGCTTCTGAAGCGCGCTTACGTCCTGGACGTAGTGCTGCT 1422
 Db 1483 ACACTAAGCAAGAGTGTATTAATCAAGTAATTTACTATCATGATGTTTGTATTAAT 1542
 Qy 1423 GCTCTGATCTGAATGCTTGGATTTGCTGTGTTTCCGCTTCTTCACTGCGCGTGC 1482
 Db 1543 GGTTCAAACCTGAATCTATTGTTTCTGATTTTTCAAAATTAACCTGCTGTCAAA 1602
 Qy 1483 TTCTCGGCGCTAGGAAGAGGCG--TCGAGCGCGAGTGTCTCCAGTAGCGCGGTT 1539
 Db 1603 TATCAGGTCTTAAGCTGCGGCAATTTCTGACGAGAAATCTGCAACACCTTAAGTC 1662
 Qy 1540 CCGGCTGCTTGTCTGACTGCTCAACAGCACTCAATCGAAGCACTGGCAATGCAAT 1599
 Db 1663 CAACATGCTTCCGCAATTTTAAACGACTTTAATAAAGATGCACTGGCGCTCAAA 1722
 Qy 1600 CGCATATGTTGGTAAAGGCTCTCGATACGCGCGCTGCAATGATTAAGGCGAGTCACT 1659
 Db 1723 ACAGTCTCAATGCTTATTTAATGACCGAGCACTCAGTTAGATCCGCGCAAGTGACC 1782
 Qy 1660 GACAAAGGCTCGATCAACGAGCGCGCTGTTTGCATGAGCGGTGCGCAAGTTGATGCG 1719
 Db 1783 GATAAAGGCAACTCAATCAAGCAAGTATTAACCAAGCTGTGCGCTTTAATGATGAG 1842
 Qy 1720 CTGTATC 1726
 Db 1843 CTTTATC 1849

RESULT 4
 US-08-403-852D-5
 ; Sequence 5, Application US/08403852D
 ; Patent No. 5891695
 ; GENERAL INFORMATION:
 ; APPLICANT: Blanc, Veronique
 ; APPLICANT: Blanche, Francis
 ; APPLICANT: Crouzet, Joel
 ; APPLICANT: Jacques, Nathalie
 ; APPLICANT: Lacroix, Patricia
 ; APPLICANT: Thibaut, Denis
 ; APPLICANT: Zagorec, Monique
 ; APPLICANT: Debussche, Laurent
 ; APPLICANT: De Crecy-Lagard, Valerie
 ; TITLE OF INVENTION: Polypeptides Involved In The
 ; TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences
 ; TITLE OF INVENTION: Coding For These Polypeptides And Their Use
 ; NUMBER OF SEQUENCES: 43
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
 ; STREET: 1300 I Street, N.W., Suite 700
 ; City: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20005-3315
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/403,852D
 ; FILING DATE: 10-MAY-1995
 ; PRIORITY APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/FR 93/00923
 ; FILING DATE: 25-SEP-1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: FR 92/11441
 ; FILING DATE: 25-SEP-1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Meyers, Kenneth J.
 ; REGISTRATION NUMBER: 25,146

REFERENCE/DOCKET NUMBER: 03806.0054-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1879 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: S.pristinaespiralis
FEATURE:
NAME/KEY: CDS
LOCATION: 110..1858
US-08-403-852D-5

Query Match 3.2%; Score 57.4; DB 2; Length 1879;
Best Local Similarity 56.7%; Pred. No. 1.9e-06;
Matches 106; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 1071 CACCGACCGCTGTCATGCTGTTACATTGGGCTGCACGCGCTGCTGCAGTCA 1130
DB 1186 CACCCAGGCGCGCGGTGTCGAGCGCCGACGAGATCGGTCGCCACCCCGACGCGCG 1245
QY 1131 GCTCGTCCGGTGCATGGGAATTGGAAGGCGCTTTCATGTCGCCACGTCATGACCG 1190
DB 1246 GCCCGTCCCGCGGTGAGACCGGTGAATGCTCACC CGCGCGCGCTTACACGCTGCGCG 1305
QY 1191 CTACTGCGCTGCTCTTGAACAAATGCCCAAGCGTTTGACGAGAGAGCTATTACTGCTC 1250
DB 1306 CTACTACCGGCGCGCGCGACGACGCGCGCGCTTACCGAGAGCGCTTCTACCGCAG 1365
QY 1251 CGGTGAT 1257
DB 1366 CGGCGAT 1372

RESULT 5
US-08-510-646B-5
Sequence 5, Application US/08510646B
Patent No. 6077699
GENERAL INFORMATION:
APPLICANT: Blanc, Veronique
APPLICANT: Blanche, Francis
APPLICANT: Crouzet, Joel
APPLICANT: Jacques, Nathalie
APPLICANT: Lacroix, Patricia
APPLICANT: Thibaut, Denis
APPLICANT: Zagorec, Monique
APPLICANT: De Crey-Lagard, Valerie
TITLE OF INVENTION: Polypeptides Involved In The
TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences
TITLE OF INVENTION: Coding For These Polypeptides And Their Use
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Parabow, Garrett & Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/510,646B

FILING DATE: 03-AUG-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,852
FILING DATE: 10-MAY-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR 93/00923
FILING DATE: 25-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 92/11441
FILING DATE: 25-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03806.0054-01000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1879 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: S.pristinaespiralis
FEATURE:
NAME/KEY: CDS
LOCATION: 110..1858
US-08-510-646B-5

Query Match 3.2%; Score 57.4; DB 3; Length 1879;
Best Local Similarity 56.7%; Pred. No. 1.9e-06;
Matches 106; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 1071 CACCGACCGCTGTCATGCTGTTACATTGGGCTGCACGCGCTGCTGCAGTCA 1130
DB 1186 CACCCAGGCGCGCGGTGTCGAGCGCCGACGAGATCGGTCGCCACCCCGACGCGCG 1245
QY 1131 GCTCGTCCGGTGCATGGGAATTGGAAGGCGCTTTCATGTCGCCACGTCATGACCG 1190
DB 1246 GCCCGTCCCGCGGTGAGACCGGTGAATGCTCACC CGCGCGCGCTTACACGCTGCGCG 1305
QY 1191 CTACTGCGCTGCTCTTGAACAAATGCCCAAGCGTTTGACGAGAGAGCTATTACTGCTC 1250
DB 1306 CTACTACCGGCGCGCGCGACGACGCGCGCGCTTACCGAGAGCGCTTCTACCGCAG 1365
QY 1251 CGGTGAT 1257
DB 1366 CGGCGAT 1372

RESULT 6
US-09-231-818-5
Sequence 5, Application US/09231818
Patent No. 6171846
GENERAL INFORMATION:
APPLICANT: Blanc, Veronique
APPLICANT: Blanche, Francis
APPLICANT: Crouzet, Joel
APPLICANT: Jacques, Nathalie
APPLICANT: Lacroix, Patricia
APPLICANT: Thibaut, Denis
APPLICANT: Zagorec, Monique
APPLICANT: De Crey-Lagard, Valerie
TITLE OF INVENTION: Polypeptides Involved In The
TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences
TITLE OF INVENTION: Coding For These Polypeptides And Their Use
NUMBER OF SEQUENCES: 43

1 CORRESPONDENCE ADDRESS:
2 ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
3 STREET: 1300 I Street, N.W., Suite 700
4 CITY: Washington
5 STATE: D.C.
6 COUNTRY: USA
7 ZIP: 20005-3315
8
9 COMPUTER READABLE FORM:
10 MEDIUM TYPE: Floppy disk
11 COMPUTER: IBM PC compatible
12 OPERATING SYSTEM: PC-DOS/MS-DOS
13 SOFTWARE: Patentin Release #1.0, Version #1.30
14
15 CURRENT APPLICATION DATA:
16 APPLICATION NUMBER: US/09/231,818
17 FILING DATE:
18 PRIORITY APPLICATION DATA:
19 APPLICATION NUMBER: US/08/403,852
20 FILING DATE: 10-MAY-1995
21 APPLICATION NUMBER: PCT/FR 93/00923
22 FILING DATE: 25-SEP-1993
23 PRIOR APPLICATION DATA:
24 APPLICATION NUMBER: FR 92/11441
25 FILING DATE: 25-SEP-1992
26 ATTORNEY/AGENT INFORMATION:
27 NAME: Meyers, Kenneth J.
28 REGISTRATION NUMBER: 25,146
29 REFERENCE/DOCKET NUMBER: 03806.0054-00000
30 TELECOMMUNICATION INFORMATION:
31 TELEPHONE: (202) 408-4000
32 TELEFAX: (202) 408-4400
33 INFORMATION FOR SEQ ID NO: 5:
34 SEQUENCE CHARACTERISTICS:
35 LENGTH: 1879 base pairs
36 TYPE: nucleic acid
37 STRANDEDNESS: double
38 TOPOLOGY: linear
39 MOLECULE TYPE: cDNA
40 HYPOTHEICAL: NO
41 ANTI-SENSE: NO
42 ORIGINAL SOURCE:
43 ORGANISM: S.pristinaespiralis
44 FEATURE:
45 NAME/KEY: CDS
46 LOCATION: 110..1858
47
48 US-09-231-818-5
49
50 Query Match 3.2%; Score 57.4; DB 3; Length 1879;
51 Best Local Similarity 56.7%; Pred. No. 1.9e-06;
52 Matches 106; Conservative 0; Mismatches 81; Indels 0; Gaps 0;
53
54 QY 1071 CACCGACCGCTGTGATGCGTTACATTGGGCTGCCAGCGCTGGCTGCGAGGTCAA 1130
55 |||||
56 Db 1186 CACCCAGGGCGCGCGGTGTCGAGCGCCAGAGATACGGTGCAGCCCGACGGCCG 1245
57 |||||
58 QY 1131 GCTCGTTCGGTGTGATGGAATTTGGAAGGCGTTTCATGTCGCACTATGACGG 1190
59 |||||
60 Db 1246 GCCCGTGGCCCGGCGTGAAGCGGTGCTACCCCGCGCCCTACACGCTGCGCGG 1305
61 |||||
62 QY 1191 CTACTGGCGTCTCTGAAACAAATGCCCAGGCTTCGAGAGAGAGGCTATTACTGCTC 1250
63 |||||
64 Db 1306 CTACTACCGGGCCCCCGAGCAACGCGCGGTTCACCGAGAGCGGCTTTACCGCAG 1365
65 |||||
66 QY 1251 CGGTGAT 1257
67 |||||
68 Db 1366 CGCGCAT 1372
69
70 RESULT 7
71 US-09-635-359B-5
72 ; Sequence 5, Application US/09635359B
73 ; Patent No. 6670157
74 ; GENERAL INFORMATION:
75 ; APPLICANT: Blanc, Veronique

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

Db 1306 CTACTACCGGGCCCCGAGACACAGCCCGCGTTACCGAGACGGCTTCTACCCGAG 1365
|||
Qy 1251 CGGTGAT 1257
|||
Db 1366 CGGCGAT 1372
|||

RESULT 8

US-09-252-991A-4762/c
; Sequence 4762, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 4762
; LENGTH: 891
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4762

Query Match 2.5%; Score 44.8; DB 4; Length 891;
Best Local Similarity 52.7%; Pred. No. 0.0057;
Matches 97; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

Qy 1081 CTGTGATGCTGTGATACATTTGGGCTGCCAGCGCTGCGAGTCAAGCTCTGTCG 1140
|||
Db 717 CGGCGCGTGTGCGCGGACGAGAGTGCGGTGCTGACCGCGGCGCGGAGTCCG 658
|||
Qy 1141 GTGCGATGGGAATTGGAGAGGCGCTTTCATGCTCCGACGTCATGAGCGGCTACTGCGCT 1200
|||
Db 657 CCGGCGAGGTGCGCGGACGAGTGCCTGCGGCTCCCTACACATCCGCGGCTACTACCGC 598
|||
Qy 1201 GCTCTGAACAATATGCCCAAGCGCTTTCGACGAGGAGGCTATTACTCTCCGATGCC 1260
|||
Db 597 CTGCGGACACACAGCGCAAGGCTTTCAGCGCGGAGCGCTTTCACCGACCGGACCGG 538
|||
Qy 1261 ATCA 1264
|||
Db 537 GTCA 534
|||

RESULT 9

US-09-252-991A-4723
; Sequence 4723, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 4723
; LENGTH: 1554
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4723

Query Match 2.5%; Score 44.8; DB 4; Length 1554;
Best Local Similarity 52.7%; Pred. No. 0.0074;
Matches 97; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

Qy 1081 CTGTGATGCTGTGATACATTTGGGCTGCCAGCGCTGCGAGTCAAGCTCTGTCG 1140
|||
Db 251 CGGCGCGTGTGCGCGGACGAGAGTGCGGTGCTGACCGCGGCGCGGAGTCCG 310
|||
Qy 1141 GTGCGATGGGAATTGGAGAGGCGCTTTCATGCTCCGACGTCATGAGCGGCTACTGCGCT 1200
|||
Db 311 CCGGCGAGGTGCGCGGACGAGTGCCTGCGGCTCCCTACACATCCGCGGCTACTACCGC 370
|||
Qy 1201 GCTCTGAACAATATGCCCAAGCGCTTTCGACGAGGAGGCTATTACTCTCCGATGCC 1260
|||
Db 371 CTGCGGACACACAGCGCAAGGCTTTCAGCGCGGAGCGCTTTCACCGACCGGACCGG 430
|||
Qy 1261 ATCA 1264
|||
Db 431 GTCA 434
|||

RESULT 10

US-09-252-991A-4682
; Sequence 4682, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 4682
; LENGTH: 1752
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4682

Query Match 2.5%; Score 44.8; DB 4; Length 1752;
Best Local Similarity 52.7%; Pred. No. 0.0078;
Matches 97; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

Qy 1081 CTGTGATGCTGTGATACATTTGGGCTGCCAGCGCTGCGAGTCAAGCTCTGTCG 1140
|||
Db 1201 CGGCGCGTGTGCGCGGACGAGAGTGCGGTGCTGACCGCGGCGCGGAGTCCG 1260
|||
Qy 1141 GTGCGATGGGAATTGGAGAGGCGCTTTCATGCTCCGACGTCATGAGCGGCTACTGCGCT 1200
|||
Db 1261 CCGGCGAGGTGCGCGGACGAGTGCCTGCGGCTCCCTACACATCCGCGGCTACTACCGC 1320
|||
Qy 1201 GCTCTGAACAATATGCCCAAGCGCTTTCGACGAGGAGGCTATTACTCTCCGATGCC 1260
|||
Db 1321 CTGCGGACACACAGCGCAAGGCTTTCAGCGCGGAGCGCTTTCACCGACCGGACCGG 1380
|||
Qy 1261 ATCA 1264
|||
Db 1381 GTCA 1384
|||

RESULT 11

US-09-535-521-1
; Sequence 1, Application US/09535521
; Patent No. 6410714
; GENERAL INFORMATION:
; APPLICANT: Weber, Eric R.
; APPLICANT: McCall, Catherine A.
; TITLE OF INVENTION: NOVEL CANINE LOW AFFINITY IGE RECEPTOR (CANINE CD23)
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES AND USES THEREOF

```
FILE REFERENCE: AL-5
CURRENT APPLICATION NUMBER: US/09/535,521
CURRENT FILING DATE: 2000-03-24
EARLIER APPLICATION NUMBER: 60/125,913
EARLIER FILING DATE: 1999-03-24
NUMBER OF SEQ ID NOS: 26
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 2851
TYPE: DNA
ORGANISM: Canis familiaris
FEATURE:
NAME/KEY: CDS
LOCATION: (199)..(1077)
US-09-535-521-1
```

```
Query Match      2.4%; Score 41.6; DB 4; Length 2851;
Best Local Similarity 55.6%; Pred. No. 0.082;
Matches 80; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
```

```
QY 664 GGTGAAGAGCGCGGCTGTGTGACCTGTTGCCGGAACACACCTTCGGGGGAGC 723
DB 976 GCGGAGACTCGCTGATGATGACGGCTCGGGCAGTGAATGACGCTTCTGCGGAGC 1035
QY 724 CACAACATCGGACATCGTTGTACAACGGCGACAGTACTTGAACGAGTAAACA 783
DB 1036 TCCTGAGAGCGTGGGTGTGACCGGCTGGCAGCTGACTGACCCACCGCGCTC 1095
QY 784 ACCGCCCAAGGTTTCCCGGAGAG 807
DB 1096 GGCACCCCGGGTTCTGCTGGAGTG 1119
```

RESULT 12

```
US-09-535-521-3/C
Sequence 3, Application US/09535521
Patent No. 6410714
GENERAL INFORMATION:
APPLICANT: Weber, Eric R.
APPLICANT: McCall, Catherine A.
TITLE OF INVENTION: NOVEL CANINE LOW AFFINITY IGE RECEPTOR (CANINE CD23)
FILE REFERENCE: AL-5
CURRENT APPLICATION NUMBER: US/09/535,521
CURRENT FILING DATE: 2000-03-24
EARLIER APPLICATION NUMBER: 60/125,913
EARLIER FILING DATE: 1999-03-24
NUMBER OF SEQ ID NOS: 26
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 3
LENGTH: 2851
TYPE: DNA
ORGANISM: Canis familiaris
US-09-535-521-3
```

```
Query Match      2.4%; Score 41.6; DB 4; Length 2851;
Best Local Similarity 55.6%; Pred. No. 0.082;
Matches 80; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
```

```
QY 664 GGTGAAGAGCGCGGCTGTGTGACCTGTTGCCGGAACACACCTTCGGGGGAGC 723
DB 1876 GGGGAGACTGCTGATGATGACAGGGCTCGGGCAGTGAATGACGCTTCTGCGGAGC 1817
QY 724 CACAACATCGGACATCGTTGTACAACGGCGACAGTACTTGAACGAGTAAACA 783
DB 1816 TCCTGAGAGCGTGGGTGTGACCGGCTGGCAGCTGACTGACCCACCGCGCTC 1757
QY 784 ACCGCCCAAGGTTTCCCGGAGAG 807
DB 1756 GGCACCCCGGGTTCTGCTGGAGTG 1733
```

RESULT 13

```
US-08-804-227C-7/C
Sequence 7, Application US/08804227C
Patent No. 5876991
GENERAL INFORMATION:
```

```
APPLICANT: Dehoff, Bradley S.
APPLICANT: Kunstos, Stuart A.
APPLICANT: Roelck, Paul R., Jr.
APPLICANT: Sutton, Kimberly L.
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSER: THOMAS G. PLANT 1501
STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
```

```
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII(DOS) Text only
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,227C
FILING DATE: February 21, 1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Plant, Thomas G.
```

```
REGISTRATION NUMBER: 35,784
REFERENCE/DOCKET NUMBER: X-8231
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-2459
```

```
INFORMATION FOR SEQ ID NO: 7:
```

```
SEQUENCE CHARACTERISTICS:
LENGTH: 4437 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
```

```
FEATURE:
NAME/KEY: CDS
LOCATION: 350..14002
```

```
FEATURE:
NAME/KEY: CDS
LOCATION: 14046..20036
```

```
FEATURE:
NAME/KEY: CDS
LOCATION: 20110..31284
```

```
FEATURE:
NAME/KEY: CDS
LOCATION: 31329..36071
```

```
FEATURE:
NAME/KEY: CDS
LOCATION: 36155..41830
US-08-804-227C-7
```

```
Query Match      2.3%; Score 41; DB 2; Length 44377;
Best Local Similarity 52.7%; Pred. No. 0.43;
Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
```

```
QY 732 CGGACATCGTGTGTACAACGGCGGACGTAATCTTGAACGAGTAAACACGCCCA 791
DB 1139 CGCGCTGTTGCGGACGCGACGCCACGACGACGAGTACAGGGGCTCCCTCGGCCA 1080
QY 792 AGGTTCCGCGGAGACGCTTCGCACTTGAGCCGAATCTCCACGCGTACTCACTG 851
DB 1079 GGGCTGCGGAGAGGCTTCAGGACGACGCGCGCGCTTACCGGACATAGCGT 1020
QY 852 GCCGAAGCTGGAGGAATAGTGGTGCCCTTGAGCAGACAGTACC 900
DB 1019 TGGAACGGGCTGCAAGTGTGTGACAGCGGCTCGGCTGACAGCGGCC 971
```

RESULT 14
US-08-804-198-1/C
Sequence 1, Application US/08804198
Patent No. 5945320
GENERAL INFORMATION:
APPLICANT: Burett, Stanley G.
APPLICANT: Kuhnke, Stuart A.
APPLICANT: Rao, Nagara R.
APPLICANT: Richardson, Mark A.
APPLICANT: Roebuck, Paul R., Jr.
TITLE OF INVENTION: PLATENOIDE SYNTHASE GENE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: PAUL R. CANTRELL, 1138
STREET, LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh 7.0
SOFTWARE: Microsoft Word 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,198
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: CANTRELL, PAUL R.
REGISTRATION NUMBER: 36,470
REFERENCE/DOCKET NUMBER: P9113
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-3885
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 44377 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 350..14002
FEATURE:
NAME/KEY: CDS
LOCATION: 14046..20036
FEATURE:
NAME/KEY: CDS
LOCATION: 20110..31284
FEATURE:
NAME/KEY: CDS
LOCATION: 31329..36071
FEATURE:
NAME/KEY: CDS
LOCATION: 36155..41830
US-08-804-198-1

Query Match 2.3%; Score 41; DB 2; Length 44377;
Best Local Similarity 52.7%; Pred. No. 0.43;
Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 732 CGGATCTGTGTGACACGCGCGGACGCTACTTGTGACGACGGTAACCAACCGCCCA 791
DB 1139 CGCGGTGTGTGCGCAGCGCGACGCGACGACGACGAGTACCGGGTCCCGTGGGCA 1080
QY 792 AGGATTCGCGAGAGCGCTTGACGCACTTGAGGCAATCTCTCCACTGCGTACTACTGT 851
DB 1079 GGGGTGCGCGAGCGGCTTCAGAGCGAGCGCGCGCGCCCTTACCGCGACATAGCCGT 1020
QY 852 GCCGAAAGGCTGGAGGAAATTAGTGGGTGCCCTTGAGCGAGACGTAAC 900
DB 1019 TGCGACGGGCGTGAAGGTGTGCGAGCGGCGGTGCGGTGACGCGGCC 971

RESULT 15
US-09-252-991A-7079/C
Sequence 7079, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT FILING DATE: 1999-02-18
CURRENT APPLICATION NUMBER: US/09/252,991A
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 7079
LENGTH: 609
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-7079

Query Match 2.2%; Score 38.4; DB 4; Length 609;
Best Local Similarity 51.3%; Pred. No. 0.34;
Matches 118; Conservative 0; Mismatches 106; Indels 6; Gaps 1;

QY 538 GCGCCGATACGATGTCGCAAGTCTTGTTCATCTTGTGCTCTACCAAACTGCTTAAGCG 597
DB 236 GTCGCGGAGACATACCTTCTTCATATACACCTCCGCTTCAACCGCATTGCCCAAGGCG 177
QY 598 GTGCCGACTACTAGCGCAATGCTCTGCGCCATCAAGATGCTTCGCAAACTTCCG 657
DB 176 GTGAGGTACCCATGCGCAACTGCTGCGCAACAGGTGCTGATCCGCGAGGCTTGCGC 117
QY 658 GTTTTGTGAAGAGCCCGGCTGCTGTGAGACTGTTGCGGTGGAACCAACCTTCGCG 717
DB 116 ATCGTGTCCGACGAC-----GTGATGCTGAGCTGGCTGCGGTGACACATAGGGC 63
QY 718 GGCAGCCACAACATGCGCATGTGTTGTAACACGCGCGCACGTAACCT 767
DB 62 CTGATCGGCGGCGCTGTGCAACGATCTTCAGCGGCGGTACCTGCGGT 13

Search completed: February 7, 2005, 21:38:26
Job time : 182 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2005, 15:41:02 ; Search time 826 Seconds

(without alignments)
11248.759 Million cell updates/sec

Title: US-09-750-986D-29

Perfect score: 1770

Sequence: 1 atgcgtctctcgcagcgcgc.....gtgacgagcgccacactgtga 1770

Scoring table: IDENTITY_NUC

Gap 10.0 , Gape 1.0

Searched: 413486 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: N_Geneseq_23Sep04:*

1: geneseqn1980s:*
2: geneseqn1980s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	445.4	25.2	1902	9 ADA31106	ADA31106 DNA encod
2	404.6	22.9	1203	3 AAA39722	AAA39722 Pseudomon
3	397.2	22.4	2188	3 AAA39720	AAA39720 Pseudomon
4	396.8	22.4	2171	3 AAA39721	AAA39721 Pseudomon
5	73	4.1	1543	3 AAA39731	AAA39731 Pseudomon
6	73	4.1	2509	3 AAA39730	AAA39730 Pseudomon
7	73	4.1	2526	3 AAA39729	AAA39729 Pseudomon
8	57.4	3.2	1879	2 AA064205	AA064205 sRNA gene
9	53.2	3.0	2000	8 ADA71938	ADA71938 Rice gene
10	46.8	2.6	536	10 AD868842	AD868842 Minorily
11	46.6	2.6	5057	2 AA223757	AA223757 P. Fluore
12	46.4	2.6	1644	8 ACA42560	ACA42560 Prokaryot
13	45.6	2.6	2000	8 ADA71938	ADA71938 Rice gene
14	45.4	2.6	5451	5 AAF26319	AAF26319 Pseudomon
15	45.4	2.6	6522	8 ACA43663	ACA43663 Prokaryot
16	44.8	2.5	891	11 AB06158	AB06158 Pseudomon
17	44.8	2.5	1554	11 AB06119	AB06119 Pseudomon
18	44.8	2.5	1752	11 ABD06078	ABD06078 Pseudomon
19	44.4	2.5	1632	8 ACA51097	ACA51097 Prokaryot
20	43.4	2.5	342	8 ACA14026	ACA14026 Prokaryot
21	43.4	2.5	410	8 ACA15140	ACA15140 Prokaryot

22	43.4	2.5	1701	8 ACA18952	ACA18952 Prokaryot
23	43	2.4	10732	3 AAA10594	AAA10594 Gene enco
24	41.6	2.4	2851	6 ABN86559	ABN86559 Canine 1D
25	41.6	2.4	2851	6 ABN86560	ABN86560 Canine 1D
26	41.4	2.3	1571	3 AAA39728	AAA39728 Pseudomon
27	41.4	2.3	2506	3 AAA39727	AAA39727 Pseudomon
28	41.4	2.3	2539	3 AAA39726	AAA39726 Pseudomon
29	41.2	2.3	9192	12 AD026464	AD026464 Brevibact
30	41	2.3	44377	2 AAT80414	AAT80414 Platanoli
31	41	2.3	44377	2 AAT80508	AAT80508 Platanoli
32	40.6	2.3	1754	4 AAF81366	AAF81366 Quorum se
33	40.6	2.3	13029	4 AAS51470	AAS51470 Pseudomon
34	40.6	2.3	13029	8 ACA19370	ACA19370 Prokaryot
35	40.2	2.3	1803	10 ACF71666	ACF71666 Phototrab
36	40.2	2.3	110000	10 ACF67367_48	ACF67367_48
37	40.2	2.3	110000	10 ACF65387_0	ACF65387 Phototrab
38	40.2	2.3	349980	6 ABO81844	ABO81844 Bifidobac
39	39.6	2.2	1668	10 AB237954	AB237954 N. gonorr
40	38.8	2.2	1744	2 AA063900	AA063900 Acyl COA
41	38.6	2.2	2003	5 AAS85442	AAS85442 DNA encod
42	38.6	2.2	3286	6 ABS51519	ABS51519 Human CDN
43	38.6	2.2	4468	6 ABS51504	ABS51504 Human CDN
44	38.6	2.2	13855	10 ADE79055	ADE79055 Human pro
45	38.4	2.2	609	11 ABD08475	ABD08475 Pseudomon

ALIGNMENTS

RESULT 1	ADA31106	standard; DNA, 1902 BP.
ID	ADA31106	
XX	ADA31106;	
AC	20-NOV-2003	(first entry)
DT		
XX	DNA encoding Acinetobacter baumannii protein #2393.	
DE	day; gene; Acinetobacter baumannii; bacterial disease; antibacterial;	
XX	vaccine; plant biocontrol agent.	
KW		
XX	Acinetobacter baumannii.	
OS		
XX	US6562958-B1.	
PN		
XX	13-MAY-2003.	
PD		
XX	04-JUN-1999;	99US-00328352.
PF		
XX	09-JUN-1998;	98US-0088701P.
PR		
XX	(GENO-) GENOME THERAPEUTICS CORP.	
PA		
XX	Breton G, Bush D;	
PI		
XX	WPI, 2003-576092/54.	
XX	P-PSDB; ADA35232.	
DR		
XX	New Acinetobacter baumannii proteins and nucleic acids, useful as reagents	
PT	for diagnosing a bacterial disease, as components of antibacterial	
PT	vaccines, as targets for antibacterial drugs, or as biocontrol agents for	
PT	plants.	
XX		
XX	Example; SEQ ID NO 2393; 328bp; English.	
PS		
XX	The invention relates to isolated Acinetobacter baumannii nucleic acids.	
CC	for diagnosing a bacterial disease, as components of antibacterial	
CC	vaccines, as targets for antibacterial drugs, to detect the presence of	
CC	A. baumannii and other Acinetobacter species in a sample, in screening	
CC	compounds for the ability to interfere with the A. baumannii life cycle	
CC	or to inhibit A. baumannii infection, and as biocontrol agents for	

CC plants. The present sequence represents DNA encoding an A. baumannii
 CC protein.

XX Sequence 1902 BP; 566 A; 420 C; 392 G; 524 T; 0 U; 0 Other;

Query Match 25.2%; Score 445.4; DB 9; Length 1902;

Best Local Similarity 55.2%; Pred. No. 1e-115;

Matches 932; Conservative 0; Mismatches 746; Indels 9; Gaps 3;

QY 46 GAGCGTCTGAGATGGGCTTAAGACCCGTCAGAACCAACCTGCTGTCAGAGGCG 105
 DB 166 GATCGTTTAATTCATTTTGCACAAACCAACCCGACCATATTTTGCAGCAAAAGCAAT 225
 QY 106 GCAAATGGGAGATGCGTCCTATCAGTACGCGAAATGTTCCACAAGTCGCGCCATC 165
 DB 226 GCTCAAGGCGAATGGGTCACAAAGTATGCAAGAGTTCTACAGCGCATGGACATT 285
 QY 166 GCACAGAGCTTGCTTCTTACGGAATTCGCGCAGAGCGTCCGCTGTTATGCTCTGA 225
 DB 286 GCTCAAGCTTTCATGCTCGTAAATTTAAGCAAGAAAGCTCTGTCATTTAAGTGT 345
 QY 226 AATGACCTGGAGATCTTACGCTGGCATTTGGGGCTATGATGCGGCAATTCCTATTCG 285
 DB 346 AATGATCTGAACTTTAATTAACATGCTGTATGAGCTGCCATGCTGGCAGCGCTTCTCG 405
 QY 286 CCGGTCTCTCTGCTTATTCATCTGCTGTCCGCAAGATTGGCGAAGCTGCTCAATGTA 345
 DB 406 GCTATTTCCCTGCTACTCTGATTTTCAAGACTTTGGCAAACTCAAACTATGTTT 465
 QY 346 GGTCTCTGCAACCGGACCTGCTTTCGTCGCGATGACAGACCTTTCCAGCGCAAT 405
 DB 466 GAAGTGTCTACACCTGGTATGTTTATGCCAGGATGACAACTTTTGCAAAGCGATT 525
 QY 406 GAGACCATTTCTGCGGAGCAGCGTCCCGCAATCTTCACTGCGGCGAAATTTGGCGGCG 465
 DB 526 CAGGCAATGATTTACGCTGTATTTGAAGTGTGACCAATTAAGGAATAGTGGCGATCAG 585
 QY 466 CGCAGCGTGAATTTGACAGCTGCTGAGCAGCTGTGGGATTTGAGCGAATTAATGCC 525
 DB 586 ATCTGACGCTTTTCAATGCTGTATGATACACAGTTTCAATGTTCAAGAGTTT --- 642
 QY 526 TTTGGCGCACTGCGCCCATAGATTTGCCAAGTTCTTGTTCATTTGCTCTACCAAA 585
 DB 643 TATCAAAACCTTATGAAACCAAGATTGCCAAATTTCTGTTTACATCAGGTTTCAACTAAA 702
 QY 586 CTGCTAAGGCGGTCGCACTACTCAGCGAATGCTGCGCAATCAGCAGATGCTTCTG 645
 DB 703 TTACTTAAGGCTGTACCGACCAACATTTAATGTTGTTGTTAATCAGCAATGTTATTTG 762
 QY 646 CAAACTTTCCCGGTTTTTGTGAAGACCGCGGTGCTGTGAGCTGTGCGGTGGAAC 705
 DB 763 CAGACTTTCCCTGAGTTTGAAGAAACACGCTGTCTACTCGACTGCTGTCTTGGCAC 822
 QY 706 CACACCTTGGCGGCGACCAACATCTGGCATCTGTGTTTGAACACGCGCGCACTACTAC 765
 DB 823 CACACATTTGGCGGCGAGTCACAATGTCGCGATCGCACTTATTAACGCGCGGATGATTTAC 882
 QY 766 CTGACGACGCGTAAACCAACCGCCCAAGGGTTGCGCGAGCGCTTGCACATTTGAGCGAA 825
 DB 883 ATTGATGATGCAAAACCGTTGCAAGAAATTTGAGCAAACTATGCTAATCTCAAGAA 942
 QY 826 ATCTCTCCACTGTACTCTACTGTGCGAAAGGCTGGAGGAATTAATGAGTGCCTT 885
 DB 943 ATTTCTCAACTGTTTATTTAATGTCGCAAAAGTTGGGAGAAACTCACCGAAGCGTTA 1002
 QY 886 GACCGAGACGTAACCTGCGCGCAACGTTCTTGTCTGCAATGAAGCTTCTTCTTGGCG 945
 DB 1003 GAAAAAGATGAAGATTTAAGAGCGCTTTTGTGCCAAAGTTAAATTTATTTCTTGGCC 1062
 QY 946 GCGGCTGTGGTTGTGCAAGGGAATCTGGGATGTTTTGACCGGGTGTGTAACGCACTGT 1005
 DB 1063 GGTGCTGCACTTTCAAGAGCGGCTGGACAGACTGATTAATTTGCTTCAGCAACTTGC 1122

QY 1006 GGTAGCGCAATTCGATGATGGCGGTCTGGGCGATGACGAGACGTCTCTTCTGCACT 1065
 DB 1123 GGAGAAAAATCCGCAATTAATGAGCGGATGGGCGATGACCGAACTCTCTCTTGTGCT 1182
 QY 1066 TTTACCAACCGGACCGGCTGTCATGCTGTGTTACATTTGGGCTGCAACGCTGCTCGAG 1125
 DB 1183 TTTTCAACTGGGCCCAAGCGTAATGCGTGTATTTGTTTACCTCTCTCGGAGTCGAA 1242
 QY 1126 GTCACTCGTTCGCGTCATGAGGAATTTGAAGGCGTTTTCCATGATTCGACGTCATG 1185
 DB 1243 ATTAAGCTATTTCCATGATGATGACCAACTGATTTGCGTTGCTGTGGCAACAGCTATG 1302
 QY 1186 AGCGCTACTGGCG---TGCCTGTAACAATAATGCCAAGCTTTCACAGAGAGCTAT 1242
 DB 1303 AAGGCTATTTGGCGCTTAAAGGCGGACCAACAAGTACTATTTTATGATGTAAGGCTTT 1362
 QY 1243 TACTGCTCGGTCATGTCATCAATTTGGCAGATCCCGATCCGATCCGAAAGGTCGATG 1302
 DB 1363 TTTCAATACAGGCGATCCGTCGTTAATGTCGATGTCATGATCAACAAAGGCTTAATG 1422
 QY 1303 TTTGACGTCGAATTCGTAAGACTTCAAGCTGTCCCTCAGGGGTATTTGTACAGCTTGGG 1362
 DB 1423 TACGACGAGAGATTCGCAAGACTTTAACTCAATACAGGCACTTTGTGATATGCGG 1482
 QY 1363 CCAATTCGCAACGCGGCGGCTTCTGAAAGCGGCTTTACGTCCTGACGATGTTGCT 1422
 DB 1483 ACACTACGCAACAAGATGCTTATTCAGGTAATTTACTATCCAGATGTTGTATTAAT 1542
 QY 1423 GCTCCGATCGTAATGCTTGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1482
 DB 1543 GGTTCAAACCTGATGCTATTTGTTTCTGATTTTTCAAAATTTAAGCCTTGTGCTCAA 1602
 QY 1483 TTTGCGGCGCTAGGAAAGAGCG--TCGACGCGGAGTGTCTTCCAGTGAAGCGGTT 1539
 DB 1603 TATGACAGCTTAAAGCTGAGCGCAATTTCTGACAGCAGATGACTGCAACCTTAAGTC 1662
 QY 1540 CCGGCGCTGTTGCTGATCTGCTCAACAGACTCAATCGAAGAGCAATGCGCAATGCAAT 1599
 DB 1663 CAACAAATGTTCCGCAATTTTAAAGACCTTTAAATTAAGATGCGCTGCGACCTCAAT 1722
 QY 1600 CGCATGATGAGGCTAGGCGCTCCCGATACGCGCGCTGATGATTAAGGCGAGTCACT 1659
 DB 1723 ACAGTCTCAATGCTTTATTTAATGACCGAGCCACTCACTTATGATGCGGCGAAGTAC 1782
 QY 1660 GACAAAGGCTCGATCAACAGCGCGCTGTTTTCGAATGCGGCTGCGCAAGTTGATGCG 1719
 DB 1783 GATTAAGGCAACCTCATTAAGCACTATTAACAAAGCTGCGCTTAAATTTGATGAG 1842
 QY 1720 CTGTATC 1726
 DB 1843 CTTTATC 1849

RESULT 2
 AAA9722
 ID AAA9722 standard; DNA; 1203 BP.
 AAA9722;
 21-SEP-2000 (first entry)

Pseudomonas sp. HR199 fcs-delta DNA.
 Eugenol; ferulic acid; conferyl alcohol; conferyl aldehyde; vanillin;
 vanillinic acid; flavouring; ferulic acid-coh-synthetase; fcs; fcs-delta;
 ds.
 Pseudomonas sp.
 DE19850242-A1.
 04-MAY-2000.

PF 31-OCT-1998; 98DE-01050242.
XX
PR 31-OCT-1998; 98DE-01050242.
XX
PA (HAAR) HAARMANN & REIMER GMBH.
XX
PI Rabenhorst J, Steinhuechel A, Priefert H, Overhage J;
XX
DR WPI; 2000-340642/30.
XX P-PSDB; AA67972.
XX
PT Transformed organisms, used for production of vanillin and other
XX methoxyphenols, have altered catabolism of eugenol or ferulic acid.
XX
PS Claim 8; Fig 2i; 80pp; German.
XX
XX This invention describes novel transformed and/or mutated uni- or multi-
CC cellular organisms (A) in which enzymes (I) involved in catabolism of
CC eugenol (E) and/or ferulic acid (FA) are inactivated in such a way that
CC the organism accumulates the intermediates conferyl alcohol (CA1),
CC conferyl aldehyde (CA2), FA, vanillin (V) and/or vanillic acid (VA).
CC INDEPENDENT CLAIMS are also included for the following: (a) gene
CC structures in which sequences that encode one or more of the enzymes CA1-
CC aldoase, CA2-dehydrogenase, V-dehydrogenase, enoyl-CoA hydratase-
CC aldolase, beta-ketochiolase, V-dehydrogenase and VA-demethylase have been
CC altered and/or inactivated; (b) vector containing at least one gene
CC structure of (a); (c) biotechnical production of organic compounds
CC (particularly alcohols, aldehydes or organic acids) using (A); and (d)
CC production method of (A). (A) are used to prepare organic compounds,
CC preferably alcohols, aldehydes and organic acids, especially CA1, CA2,
CC FA, V and/or VA, most particularly V, a well-known natural flavouring.
CC The inexpensive starting material eugenol can be converted to V in a
CC single step. This sequence encodes the Pseudomonas sp. HR199 ferulic acid
CC -CoA-synthetase, fcs-delta which is described in the method of the
CC invention
XX
XX Sequence 1203 BP; 217 A; 337 C; 357 G; 292 T; 0 U; 0 Other;
SQ
Query Match 22.9%; Score 404.6; DB 3; Length 1203;
Best Local Similarity 91.6%; Pred. No. 3,4e-104;
Matches 428; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
QY 1 ATGCGTTCTCTCGAGGCGCTTCTCCCTCCGCGGTGGAATTCTTGAGCGTTCGAGCAT 60
DB 342 ATGCGTTCTCTCGAGGCGCTTCTCCCTCCGCGGTGGAATTCTTGAGCGTTCGAGCAT 401
QY 61 TGGGCTTAAGACCCGTCAGAACAAACCTGCGTTGCTCCAGGCGGCAATGGGGAATGG 120
DB 402 TGGGCTTAAGACCCGTCAGAACAAACCTGCGTTGCTCCAGGCGGCAATGGGGAATGG 461
QY 121 CGTCGTATCAGCTACCGGAAATGTTCCACACGTCGCGGCATCGACAGAGTTGCTT 180
DB 462 CGTCGTATCAGCTACCGGAAATGTTCCACACGTCGCGGCATCGACAGAGTTGCTT 521
QY 181 CTTTACGCAATATCGAGAGCGCTCGCTGTTATGCTCTTGGAATGACCTGGAACAT 240
DB 522 CTTTACGCAATATCGAGAGCGCTCGCTGTTATGCTCTTGGAATGACCTGGAACAT 581
QY 241 CTTTACGTCGCAATTTGGGGCTATGTATGCGGCAATTCCTATTCGCCGATGTCCTGCT 300
DB 582 CTTTACGTCGCAATTTGGGGCTATGTATGCGGCAATTCCTATTCGCCGATGTCCTGCT 641
QY 301 TATTCATGCTGTGCGCAAGATTGGCGAAGCTGTGTCACATCGTAGTCTTTCGCAACG 360
DB 642 TATTCATGCTGTGCGCAAGATTGGCGAAGCTGTGTCACATCGTAGTCTTTCGCAACG 701
QY 361 GAGATGCTCTTGTGCGCAATGACAGACCTTTCAGGCGGCAATGAGACCATCTTCCG 420
DB 702 GAGATGCTCTTGTGCGCAATGACAGACCTTTCAGGCGGCAATGAGACCATCTTCCG 761
QY 421 GAGATGCTCTTGTGCGCAATGACAGACCTTTCAGGCGGCAATGAGACCATCTTCCG 467
DB 762 GAGATGCTCTTGTGCGCAATGACAGACCTTTCAGGCGGCAATGAGACCATCTTCCG 808

RESULT 3
AAA39720
ID AAA39720 standard; DNA; 2188 BP.
XX
XX AAA39720;
AC
XX 21-SEP-2000 (first entry)
DT
XX
XX Pseudomonas sp. HR199 fcs-omega-Km DNA.
DE
XX Eugenol; ferulic acid; conferyl alcohol; conferyl aldehyde; vanillin;
KM vanillic acid; flavouring; ferulic acid-CoA-synthetase; fcs;
KM fcs-omega-Km; de.
XX
XX Pseudomonas sp.
OS
XX DE19850242-A1.
XX PN
XX 04-MAY-2000.
PD
XX
XX 31-OCT-1998; 98DE-01050242.
PP
XX 31-OCT-1998; 98DE-01050242.
PR 31-OCT-1998; 98DE-01050242.
XX
XX (HAAR) HAARMANN & REIMER GMBH.
PA
PI Rabenhorst J, Steinhuechel A, Priefert H, Overhage J;
XX
XX WPI; 2000-340642/30.
DR P-PSDB; AA67970.
XX
XX Transformed organisms, used for production of vanillin and other
PT methoxyphenols, have altered catabolism of eugenol or ferulic acid.
XX
XX Claim 8; Fig 2g; 80pp; German.
PS
XX
XX This invention describes novel transformed and/or mutated uni- or multi-
CC cellular organisms (A) in which enzymes (I) involved in catabolism of
CC eugenol (E) and/or ferulic acid (FA) are inactivated in such a way that
CC the organism accumulates the intermediates conferyl alcohol (CA1),
CC conferyl aldehyde (CA2), FA, vanillin (V) and/or vanillic acid (VA).
CC INDEPENDENT CLAIMS are also included for the following: (a) gene
CC structures in which sequences that encode one or more of the enzymes CA1-
CC aldoase, CA2-dehydrogenase, V-dehydrogenase, enoyl-CoA hydratase-
CC aldolase, beta-ketochiolase, V-dehydrogenase and VA-demethylase have been
CC altered and/or inactivated; (b) vector containing at least one gene
CC structure of (a); (c) biotechnical production of organic compounds
CC (particularly alcohols, aldehydes or organic acids) using (A); and (d)
CC production method of (A). (A) are used to prepare organic compounds,
CC preferably alcohols, aldehydes and organic acids, especially CA1, CA2,
CC FA, V and/or VA, most particularly V, a well-known natural flavouring.
CC The inexpensive starting material eugenol can be converted to V in a
CC single step. This sequence encodes the Pseudomonas sp. HR199 ferulic acid
CC -CoA-synthetase, fcs-omega-Km which is described in the method of the
CC invention
XX
XX Sequence 2188 BP; 408 A; 607 C; 670 G; 503 T; 0 U; 0 Other;
SQ
Query Match 22.4%; Score 397.2; DB 3; Length 2188;
Best Local Similarity 99.3%; Pred. No. 5,6e-102;
Matches 399; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ATGCGTTCTCTCGAGGCGCTTCTCCCTCCGCGGTGGAATTCTTGAGCGTTCGAGCAT 60
DB 342 ATGCGTTCTCTCGAGGCGCTTCTCCCTCCGCGGTGGAATTCTTGAGCGTTCGAGCAT 401
QY 61 TGGGCTTAAGACCCGTCAGAACAAACCTGCGTTGCTCCAGGCGGCAATGGGGAATGG 120
DB 402 TGGGCTTAAGACCCGTCAGAACAAACCTGCGTTGCTCCAGGCGGCAATGGGGAATGG 461
QY 121 CGTCGTATCAGCTACCGGAAATGTTCCACACGTCGCGGCATTCGACAGAGCTTCTGCTT 180

```

Db      462 CGTCGTATCAGCTACGCGGAAATGTTCCAAACGTCGCGCCATCGCACAGACTTGCTT
Qy      181 CCTTACGAGCTATCGGAGAGCGCTCGCTTATGCTCTCTGGAATGACCTGGAACAT
Db      522 CTTACGAGCTATCGGAGAGCGCTCGCTTATGCTCTCTGGAATGACCTGGAACAT
Qy      241 CTTACGAGCTATCGGAGAGCGCTTATGCTCTCTGGAATGACCTGGAACAT
Db      582 CTTACGAGCTATCGGAGAGCGCTTATGCTCTCTGGAATGACCTGGAACAT
Qy      301 TATTCACTGCTGTCGCAAGATTGGGAGAGCTGCTCAATGCTGCTTCTGCAACCG
Db      642 TATTCACTGCTGTCGCAAGATTGGGAGAGCTGCTCAATGCTGCTTCTGCAACCG
Qy      361 GGACTGCTCTTGTGCTGCGGATGAGACACTTTCAGCGCGCA
Db      702 GGACTGCTCTTGTGCTGCGGATGAGACACTTTCAGCGCGCA

```

RESULT 4

AAA39721 standard; DNA; 2171 BP.

AAA39721;

21-SEP-2000 (first entry)

Pseudomonas sp. HR199 fcs-omega-Gm DNA.

Eugenol; ferulic acid; coniferyl alcohol; coniferyl aldehyde; vanillin;
vanillinic acid; flavouring; ferulic acid-CoA-synthetase; fcs;
fcs-omega-Gm; ds.

Pseudomonas sp.

DE19850242-A1.

04-MAY-2000.

31-OCT-1998; 98DE-01050242.

31-OCT-1998; 98DE-01050242.

(HAAR) HAARMANN & REIMER GMBH.

Rabenhorst J, Steinhuechel A, Priefert H, Overhage J;

WPI; 2000-340642/30.

P-PsDB; AAY87971.

Transformed organisms, used for production of vanillin and other
methoxyphenols, have altered catabolism of eugenol or ferulic acid.

Claim 8; Fig 2h; 80pp; German.

This invention describes novel transformed and/or mutated uni- or multi-
cellular organisms (A) in which enzymes (I) involved in catabolism of
eugenol (E) and/or ferulic acid (FA) are inactivated in such a way that
the organism accumulates the intermediates coniferyl alcohol (CAI),
coniferyl aldehyde (CA2), FA, vanillin (V) and/or vanillinic acid (VA).
INDEPENDENT CLAIMS are also included for the following: (a) gene
structures in which sequences that encode one or more of the enzymes CAI-
dehydrogenase, CA2-dehydrogenase, FA-CoA-synthetase, enoyl-CoA hydratase-
aldolase, beta-ketothiolase, V-dehydrogenase and VA-demethylase have been
altered and/or inactivated; (b) vector containing at least one gene
structure of (a); (c) biotechnical production of organic compounds
(particularly alcohols, aldehydes or organic acids) using (A); and (d)
production method of (A). (A) are used to prepare organic compounds,
preferably alcohols, aldehydes and organic acids, especially CAI, CA2,
FA, V and/or VA, most particularly V, a well-known natural flavouring.
The inexpensive starting material eugenol can be converted to V in a
single step. This sequence encodes the *Pseudomonas* sp. HR199 ferulic acid

CC -CoA-synthetase, fcs-omega-Gm which is described in the method of the
invention
XX
SQ Sequence 2171 BP; 443 A; 591 C; 616 G; 521 T; 0 U; 0 Other;

Query Match 22.4%; Score 396.8; DB 3; Length 2171;
Best Local Similarity 99.5%; Pred. No. 7, 2e-102;
Matches 398; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

Qy      1 ATGCGTCTCTGAGGCGCTTCCCTCCCGGGTGAATTCTGAGCGTCCAGCAT
Db      342 ATGCGTCTCTGAGGCGCTTCCCTCCCGGGTGAATTCTGAGCGTCCAGCAT
Qy      61 TGGGCTAAGACCCGTCAGAAACAACTGCTGCTGCGAGGCGGCAATGGGAATGG
Db      402 TGGGCTAAGACCCGTCAGAAACAACTGCTGCTGCGAGGCGGCAATGGGAATGG
Qy      121 CGTCGTATCAGCTACCGGAAATGTTCCACACAGTCCGGCCATGCGACAGACTTGCTT
Db      462 CGTCGTATCAGCTACCGGAAATGTTCCACACAGTCCGGCCATGCGACAGACTTGCTT
Qy      181 CCTTACGAGCTATCGGAGAGCGCTCGCTTATGCTCTGGAATGACCTGGAACAT
Db      522 CTTACGAGCTATCGGAGAGCGCTCGCTTATGCTCTGGAATGACCTGGAACAT
Qy      241 CTTACGAGCTATCGGAGAGCGCTTATGCTCTGGAATGACCTGGAACAT
Db      582 CTTACGAGCTATCGGAGAGCGCTTATGCTCTGGAATGACCTGGAACAT
Qy      301 TATTCACTGCTGTCGCAAGATTGGGAGAGCTGCTCAATGCTGCTTCTGCAACCG
Db      642 TATTCACTGCTGTCGCAAGATTGGGAGAGCTGCTCAATGCTGCTTCTGCAACCG
Qy      361 GGACTGCTCTTGTGCTGCGGATGAGACACTTTCAGCGCGCA
Db      702 GGACTGCTCTTGTGCTGCGGATGAGACACTTTCAGCGCGCA

```

RESULT 5

AAA39731 standard; DNA; 1543 BP.

AAA39731;

21-SEP-2000 (first entry)

Pseudomonas sp. HR199 vdh-delta DNA.

Eugenol; ferulic acid; coniferyl alcohol; coniferyl aldehyde; vanillin;
vanillinic acid; flavouring; vanillin dehydrogenase; vdh; vdh-delta; ds.

Pseudomonas sp.

DE19850242-A1.

04-MAY-2000.

31-OCT-1998; 98DE-01050242.

31-OCT-1998; 98DE-01050242.

(HAAR) HAARMANN & REIMER GMBH.

Rabenhorst J, Steinhuechel A, Priefert H, Overhage J;

WPI; 2000-340642/30.

P-PsDB; AAY87981.

Transformed organisms, used for production of vanillin and other
methoxyphenols, have altered catabolism of eugenol or ferulic acid.

Claim 8; Fig 2r; 80pp; German.

CC This invention describes novel transformed and/or mutated uni- or multi-cellular organisms (A) in which enzymes (I) involved in catabolism of eugenol (E) and/or ferulic acid (FA) are inactivated in such a way that the organism accumulates the intermediates conferyl alcohol (CA1), conferyl aldehyde (CA2), FA, vanillin (V) and/or vanillic acid (VA). CC INDEPENDENT CLAIMS are also included for the following: (a) gene structures in which sequences that encode one or more of the enzymes CA1-dehydrogenase, CA2-dehydrogenase, FA-CoA-synthetase, enoyl-CoA hydratase-aldolase, beta-ketochiolase, V-dehydrogenase and VA-demethylase have been altered and/or inactivated; (b) vector containing at least one gene structure of (a); (c) biotechnical production of organic compounds, particularly alcohols, aldehydes or organic acids using (A); and (d) production method of (A). (A) are used to prepare organic compounds, preferably alcohols, aldehydes and organic acids, especially CA1, CA2, FA, V and/or VA, most particularly V, a well-known natural flavouring. CC The inexpensive starting material eugenol can be converted to V in a single step. This sequence encodes the Pseudomonas sp. HR199 vanillin dehydrogenase, vdh-delta which is described in the method of the invention

SO Sequence 1543 BP; 306 A; 387 C; 473 G; 377 T; 0 U; 0 Other;

Query Match 4.1%; Score 73; DB 3; Length 1543;

Best Local Similarity 100.0%; Pred. No. 6.7e-10;

Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1698 GCGGTCGCGCAAAAGTTGATCGCTGTATCGTGTGAAGATCAATCCATGCTGCTGACGA 1757

DB 8 GCGGTCGCGCAAAAGTTGATCGCTGTATCGTGTGAAGATCAATCCATGCTGCTGACGA 67

OY 1758 GGGCACACTGTGA 1770

DB 68 GGGCACACTGTGA 80

RESULT 6

AAA39730 standard; DNA; 2509 BP.

AC AAA39730;

DT 21-SEP-2000 (first entry)

DE Pseudomonas sp. HR199 vdh-omega-Gm DNA.

KW Eugenol; ferulic acid; conferyl alcohol; conferyl aldehyde; vanillin;

KM vanillic acid; flavouring; vanillin dehydrogenase; vdh; vdh-omega-Gm; de.

OS Pseudomonas sp.

PN DE19850242-A1.

PD 04-MAY-2000.

PF 31-OCT-1998; 98DE-01050242.

PR 31-OCT-1998; 98DE-01050242.

PA (HAAR) HAARMANN & REIMER GMBH.

PI Rebenhorst J, Steinbuechel A, Priefert H, Overhage J;

DR WPI; 2000-340642/30.

DR P-PSDB; AA87980.

PT Transformed organisms, used for production of vanillin and other methoxyphenols, have altered catabolism of eugenol or ferulic acid.

PS Claim 8; Fig 2q; 80pp; German.

CC This invention describes novel transformed and/or mutated uni- or multi-cellular organisms (A) in which enzymes (I) involved in catabolism of

CC eugenol (E) and/or ferulic acid (FA) are inactivated in such a way that the organism accumulates the intermediates conferyl alcohol (CA1), conferyl aldehyde (CA2), FA, vanillin (V) and/or vanillic acid (VA). CC INDEPENDENT CLAIMS are also included for the following: (a) gene structures in which sequences that encode one or more of the enzymes CA1-dehydrogenase, CA2-dehydrogenase, FA-CoA-synthetase, enoyl-CoA hydratase-aldolase, beta-ketochiolase, V-dehydrogenase and VA-demethylase have been altered and/or inactivated; (b) vector containing at least one gene structure of (a); (c) biotechnical production of organic compounds, particularly alcohols, aldehydes or organic acids using (A); and (d) production method of (A). (A) are used to prepare organic compounds, preferably alcohols, aldehydes and organic acids, especially CA1, CA2, FA, V and/or VA, most particularly V, a well-known natural flavouring. CC The inexpensive starting material eugenol can be converted to V in a single step. This sequence encodes the Pseudomonas sp. HR199 vanillin dehydrogenase, vdh-omega-Gm which is described in the method of the invention

SO Sequence 2509 BP; 532 A; 640 C; 731 G; 606 T; 0 U; 0 Other;

Query Match 4.1%; Score 73; DB 3; Length 2509;

Best Local Similarity 100.0%; Pred. No. 8.2e-10;

Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1698 GCGGTCGCGCAAAAGTTGATCGCTGTATCGTGTGAAGATCAATCCATGCTGCTGACGA 1757

DB 8 GCGGTCGCGCAAAAGTTGATCGCTGTATCGTGTGAAGATCAATCCATGCTGCTGACGA 67

OY 1758 GGGCACACTGTGA 1770

DB 68 GGGCACACTGTGA 80

RESULT 7

AAA39729 standard; DNA; 2526 BP.

AC AAA39729;

DT 21-SEP-2000 (first entry)

DE Pseudomonas sp. HR199 vdh-omega-Km DNA.

KW Eugenol; ferulic acid; conferyl alcohol; conferyl aldehyde; vanillin;

KM vanillic acid; flavouring; vanillin dehydrogenase; vdh; vdh-omega-Km; de.

OS Pseudomonas sp.

PN DE19850242-A1.

PD 04-MAY-2000.

PF 31-OCT-1998; 98DE-01050242.

PR 31-OCT-1998; 98DE-01050242.

PA (HAAR) HAARMANN & REIMER GMBH.

PI Rebenhorst J, Steinbuechel A, Priefert H, Overhage J;

DR WPI; 2000-340642/30.

DR P-PSDB; AA87979.

PT Transformed organisms, used for production of vanillin and other methoxyphenols, have altered catabolism of eugenol or ferulic acid.

PS Claim 8; Fig 2p; 80pp; German.

CC This invention describes novel transformed and/or mutated uni- or multi-cellular organisms (A) in which enzymes (I) involved in catabolism of eugenol (E) and/or ferulic acid (FA) are inactivated in such a way that the organism accumulates the intermediates conferyl alcohol (CA1),

CC conferyl aldehyde (CA2), FA, vanillin (V) and/or vanillinic acid (VA).
 CC INDEPENDENT CLAIMS are also included for the following: (a) gene
 CC structures in which sequences that encode one or more of the enzymes CA1-
 CC dehydrogenase, CA2-dehydrogenase, FA-CoA-synthetase, enoyl-CoA hydratase-
 CC aldolase, beta-ketothiolase, V-dehydrogenase and VA-demethylase have been
 CC altered and/or inactivated; (b) vector containing at least one gene
 CC structure of (a); (c) biotechnical production of organic compounds
 CC (particularly alcohols, aldehydes or organic acids) using (A); and (d)
 CC production method of (A). (A) are used to prepare organic compounds,
 CC preferably alcohols, aldehydes and organic acids, especially CA1, CA2,
 CC FA, V and/or VA, most particularly V, a well-known natural flavouring.
 CC The inexpensive starting material eugenol can be converted to V in a
 CC single step. This sequence encodes the pseudomonas sp. HR199 vanillin
 CC dehydrogenase, vdh-omega-Km which is described in the method of the
 CC invention

SO Sequence 2526 BP; 497 A; 656 C; 785 G; 588 T; 0 U; 0 Other;
 Query Match 4.1%; Score 73; DB 3; Length 2526;
 Best Local Similarity 100.0%; Pred. No. 8.2e-10;
 Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1698 GCGGTGGCGCAAGTGCATGCGTGTGCGTGAAGATCATCCATGCTGGGTACGA 1757
 Db 8 GCGGTGGCGCAAGTGCATGCGTGTGCGTGAAGATCATCCATGCTGGGTACGA 67

QY 1758 GGGCACACTGTGA 1770
 Db 68 GGGCACACTGTGA 80

RESULT 8
 ID AA064205
 XX AA064205 standard; cDNA; 1879 BP.

AC AA064205;
 XX
 DT 25-MAR-2003 (revised)
 DT 18-NOV-1994 (first entry)
 XX

DE snbA gene encoding enzyme in streptogramin biosynthetic pathway.
 XX
 KW Antibiotic; streptogramin; snbA; snab; snac; biosynthesis; enzyme;
 KW biosynthetic pathway; Streptomyces pristinaespiralis; snbA; snbR; ds.
 XX
 OS Streptomyces pristinaespiralis.
 XX

XX Key Location/Qualifiers
 FT CDS 110..1858
 FT /*tag= a

XX FR2696189-A1.
 XX
 PD 01-APR-1994.
 XX
 PF 25-SEP-1992; 92FR-00011441.
 XX
 PR 25-SEP-1992; 92FR-00011441.
 XX
 PA (RHON) RHONE POULENC RORER SA.
 XX
 PI Blanc V, Blanche F, Crouzet J, Jacques N, Lacroix P, Thibaut D;
 PI Zagorec M;
 XX
 DR WPI; 1994-128286/16.
 DR P-PsDB; AARS4205.
 XX
 PT DNA involved in streptogramin antibiotic biosynthesis - for prodn. or bio
 PT -conversion of streptogramin(s) or prodn. of streptogramin intermediates,
 PT derivs. or hybrid antibiotics.
 XX
 PS Claim 2; Page 57-60; 83pp; French.
 XX

CC The snbA gene product is involved in the biosynthesis of streptogramins,
 CC antibiotics active against Gram-positive bacteria. The identification of
 CC the sequences encoding the enzymes involved in the biosynthetic pathway
 CC means that they can be isolated and manipulated. Mutant microorganisms in
 CC which a step in the streptogramin biosynthetic pathway is blocked can be
 CC cultured to produce streptogramin intermediates, which may later be
 CC converted to streptogramin derivatives. Recombinant cells may also be
 CC used for the bioconversion of streptogramins from one form to another or
 CC for the production of hybrid antibiotics. (Updated on 25-MAR-2003 to
 CC correct PN field.)

XX
 SQ Sequence 1879 BP; 235 A; 682 C; 720 G; 242 T; 0 U; 0 Other;
 Query Match 3.2%; Score 57.4; DB 2; Length 1879;
 Best Local Similarity 56.7%; Pred. No. 2e-05;
 Matches 106; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 1071 CACCAGACCGCTGTGCATGGCTGTGATCAATTGGGCTGCCAGCGCTGGCTGCAGATCAA 1130
 Db 1186 CACCAGGCGCGCGCTGTGCCAGGCCGACGATACCGCTGCCAGCCGACGCGCG 1245

QY 1131 GCTGTTCCGATGATGGGAAATTGGAAGGCGCTTCCATGTCGCCACGTCATGAGCGG 1190
 Db 1246 GCGCGTCCCGCGCGGCGGTAAGCCGTAACCTCTACCGCGGCCCTTACAGCTGGCGG 1305

QY 1191 CTACTGCGGTCTCTCTGAACAAATGCCCAAGCGTTGACGAGAAAGCTATTACTGCTC 1250
 Db 1306 CTACTACCGGCGCGCGGACGACAAAGCCGCGGCTTACCGAGAGCGGCTTACCGCAG 1365

QY 1251 CGGTGAT 1257
 Db 1366 CGCGGAT 1372

RESULT 9
 ID ADA71938
 XX ADA71938 standard; DNA; 2000 BP.

AC ADA71938;
 XX
 DT 20-NOV-2003 (first entry)
 DT
 XX

DE Rice gene, SEQ ID 5263.
 XX
 KW Plant; bacterial infection; fungal infection; viral infection; rice;
 KW gene; ds.
 XX
 OS Oryza sativa.
 XX
 PN WO2003000898-A1.
 XX
 PD 03-JAN-2003.
 XX
 PF 22-JUN-2001; 2001WO-IB001105.
 XX
 PR 22-JUN-2001; 2001WO-IB001105.
 XX
 PA (SYGN) SYNGENTA PARTICIPATIONS AG.
 XX
 PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
 PI Katagiri F, Qian S, Tao Y, Whitlam S, Xie Z, Zhu T, Zou G;
 XX
 DR WPI; 2003-175290/17.
 XX
 PT Identifying at least one gene involved in plant resistance or response to
 PT pathogenic infection for conferring resistance or tolerance to a plant to
 PT bacterial, fungal or viral infection by determining or detecting plant
 PT gene expression.
 XX
 PS Claim 27; SEQ ID NO 5263; 899pp; English.
 XX
 CC The present invention relates to a method (M1) for identifying genes
 CC involved in plant resistance or response to pathogenic infection. M1

comprises identifying a gene whose expression is significantly altered in the incompatible interaction of plant gene expression relative to the expression of the gene in an uninfected plant, in a mutant plant that does not express a gene associated with response to pathogenic infection, or in a corresponding incompatible or compatible interaction. (M) is useful for conferring resistance to resistance or tolerance to a plant to bacterial, fungal or viral infection. The present sequence was used to illustrate the invention.

SQ Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;

Query Match	3.0%	Score 53.2	DB 8	Length 2000
-------------	------	------------	------	-------------

Best Local Similarity 8.9%; Pred. No. 0.00032;
Matches 66; Conservative 337; Mismatches 335; Indels 4; Gaps 1.

Qy	886	GAGCGAGACACTACCCTCCCGGAAACGCTTCTTCGTCGACGAACGCTTCTTCTTCGCG	945
Db	18	RRRYMAAMMSCARGGSSRMSRKNQGSMSKRYKCCSGCKMTTRKRSKMYASASGCTG	77
Qy	946	GCGCGTGGGTGTGCGCAAGGATCTGGGATCGTTTGACCGGGTCTGTAACAGCACTGT	1005
Db	78	SKMSGGSTSGGGMCKKRYRSKRMGRGRGRMRBSMRMMGTYRRCARBGRMAAGSGRMG	137
Qy	1006	GGTAGGCGCAATTCGATGATGAGCGGCTTGAGGATGACGAGACTGCTCTTCTGCACT	1065
Db	138	GKSRMSYMMVCYARCGCSCKRKKSKGSGMGKTCRGAAGGSGWSGAKYKSGSMSKRAMM	197
Qy	1066	TTTACCAACCGGACCGCTGTCGATGCTGTGTATCAATTGGGCTGCGACGCCCTGCTCGCAG	1125
Db	198	SSCGRSRGGRRSASRYRGTGRKGTGYKMTYVSASRMCRAYMNTTSYMAOSSYTWCKSK	257
Qy	1126	GTCAGAGCTCGTTCGCGTGATGGGAAATTGGAAGGCGCTTTCAGATGCTCCGACGTCATG	1185
Db	258	RRSMWKKMKMRKRSRYGWTSMSTYKMMCTAYKKSYSRMCYMTGGGMRGATRYMRG	317
Qy	1186	AGCGGCTACTGAGCGTCTCTGAACAAATGCCAACGCCAGCTTGACAGGAAGGCTATTAC	1245
Db	318	YMSRAAMYKQWYVRYGKGMKRGWMAGRMMBSRMSKACVYWRBMMHTRRRWA	377
Qy	1246	TGCTCCGGTGATGCCATCAAAATTGGCAGATCTCCGATCTCTAGAAAGTCTGATGTT	1305
Db	378	KSRSTSRKKKMKCMRKRYKRMRYSMRSCKPRAMMKRSGAMVMDGRGCTCRMK	437
Qy	1306	GACGGTCAATTGCTGAAGACTTCAAGCTGTCCTAGGGGATTTGTACGGTTGGGCA	1365
Db	438	SYGMRRWKSMMRMSKTYKMSBMYTRMRKKKCSRTTMMKCTGTGMMGTGRCRKYSKSMK	497
Qy	1366	TTGGCGACGCGGCGGTTCTGGAAGGCGGCTCTTACGTCGTCGACGTAGTGGTTGCTGCT	1425
Db	498	RKCRRRRRGRMYRMRKRYMSARITYMRYCAARKYSYSAARKARCMYWGKXYVAMGMMK	557
Qy	1426	CCTGATGCTGAATGCTTGATGCTGCTGTTTCCGCGCTCTTCTGCACTCCCGGCTTG	1485
Db	558	RYKRYMYMKMMYMKRYKSKSWYCKMSYVASCMKSAKAKAMCKRSMAMWSKSM---	614
Qy	1486	TCGGGGCTAGAGAAAAGAGCGCTCGGACGCCAGGCTTGCCAGTGAACCGCGTTCCGGCC	1545
Db	615	-RSSRKKCKKASKSSARIVAMGGMTSGSRMSRKSTYCTWRKMGSKSTCTMYMYMSK	673
Qy	1546	TGTTTGTGCTAGCTGCTCAACAGACTCATCGAGAGAACTGGCAATGACAGTGCATC	1605
Db	674	YTYAKYSYWMYRIRYAKMCMYMRWYVYVYRSMTYMAVYTSRBMAMGMKYSGRYWS	733
Qy	1606	ATGTRGGTAGGCTCTTCGATA	1627
Db	734	YKYCCSKMSKMSYMSYMSMMA	755

```

RESULT 10
ADB68842
ID      ADB68842 standard; DNA; 536 BP.
XX
AC      ADB68842;

```

XX	DT	04-DEC-2003	(first entry)	
XX	DE	Minority luxI consensus sequence DNA 14.		
XX	KX	quorum sensing; lux homologue; luxI; ds.		
XX	OS	Unidentified.		
XX	PN	WO2003057902-A2.		
XX	PD	17-JUL-2003.		
XX	PF	08-JAN-2003; 2003WO-US000479.		
XX	PR	08-JAN-2002; 2002US-0346531P.		
XX	PT	07-JAN-2003; 2003US-00338110.		
XX	PA	(FRAU) FRAUNHOFER USA INC.		
XX	PI	Fuhrmann JJ, Romesee JA;		
XX	DR	WPI, 2003-618102/58.		
XX	PT	Detecting quorum sensing potential of a Gram-negative bacterium in a sample comprises performing a polymerase chain reaction using nucleic acid extracted from a sample containing a microorganism.		
XX	PS	Disclosure; Fig 10; 86pp; English.		
XX	CC	The invention relates to a novel method for detecting the quorum sensing potential of a microorganism in a sample which comprises performing PCR using nucleic acids extracted from a sample containing at least one type of microorganism. The method may be useful for detecting the quorum sensing potential of a microorganism in a sample by amplifying a fragment of a lux gene or homologue. The current sequence is that of the minority luxI consensus sequence DNA of the invention.		
XX	SQ	Sequence 536 BP; 20 A; 52 C; 50 G; 40 T; 0 U; 374 Other;		
QY	Query Match	2.6%; Score 46.8; DB 10; Length 536;		
QY	Best Local Similarity	11.7%; Pred. No. 0.012;		
QY	Matches	54; Conservative 220; Mismatches 187; Indels 0; Gaps 0;		
DB		382 GCAGACCTTTCCAGCGCGCAATTGAGACCATTCGCGCGAGCAGTCCCGCAATCTTC 441		
DB		73 SVRSIBYVWVSKNVVVDMMBSBKVGARFKBAGYCAKRTWCGAYMYMSAHRCSSKC 132		
QY		442 ACTGAGCGCAATTGCGCGGCGCGCGCAGCGTGAATTGACAGCTTGCGAGCAGCTT 501		
DB		133 TACVTSBTHSGSHKVRGMVSVYVBDNATSKYGGHTGYGSCSGCTGHTCCBACXCV 192		
QY		502 GGTGGGATTTGAGCGCAGATATGCTTTGCGGCAACTGGCCCCGATAGCATTTGCCAAGTTC 561		
DB		193 BSSGCBMAYCTGCTBRARAKNVSTTTKSSBMVYVSBKCBVSRVWMMYRCGSMVSVAT 252		
QY		562 TTGTTCATCTTGCGCTACCAAACTGCTTAAGCGGCGCGCACTACCTACGAGATGTC 621		
DB		253 CCVKMSRINTTGGARHTWKCAGMTWCRSCSMCRCHSGSAMVKGCGSTGVNKKRTY 312		
QY		622 TGGCGCATCAGCAGATGCTTCTGCAAACTTCCCGGTTTTTGTGTAAGAGCGCGCGGTG 681		
DB		313 TSKSPDVWBCSTSVARWSMGCSPSYVSVBSGGGCVVRKVBVNTSRTSGSVRSYSW 372		
QY		682 CTGTGATGCTGTTGCGCGTGGAAACACACCTTGGCGGCGCACCAACATCGCATCTGTG 741		
DB		373 HCRBRSRSATGAGWVCGSYWKTWCVDNMKSAAVCGGBRTVMDBSHSANCGVYVGGVYCGS 432		
QY		742 TTGACACAGCGCGGACGCTATACCTTGCACACGCTAAACACGCGCAAGGTTGCGCC 801		
DB		433 YBNHSVMNHVBRBVRDMHRYTSTSGCSYTYGGACVCRVADCRMDSCVYVMSVHBY 492		
QY		802 GAGACGCTTCGCAACTTGAGGAAATCTCTCCACTGCGTA 842		

Db 493 KYWSCKYRSMHYCVNRHBBGVNMYKXBVVGVDRBYKV 533

RESULT 11

AA223757
ID AA223757 standard; DNA; 5057 BP.

XX AA223757;

XX 27-AUG-2003 (revised)

DT 14-JUN-2000 (first entry)

XX P. fluorescens DNA encoding ICS orfA and orfD.

XX Isochorismate synthase; ICS; plant; pathogen resistance; chitinase;
KM pathogen inducible promoter; antipathogenic protein; toxin;
KM antifungal protein; albumin-type protein; hypersensitive response;
KM isochorismate pyruvate synthase; ss.

XX Pseudomonas fluorescens.

XX Key Location/Qualifiers

FT CDS 207..1382

FT /*tag= a /product= "isochorismate synthase orfA"

FT 4516..4851

FT /*tag= b

FT /product= "isochorismate pyruvate lyase orfD"

XX MO9950423-A2.

XX 07-OCT-1999.

XX 25-MAR-1999; 99MO-EP002176.

XX 31-MAR-1998; 98US-0080203P.

XX 03-APR-1998; 98US-0080625P.

XX (MOGE-) MOGEN INT NV.

XX (UYLE-) RIJCKSUNITV LEIDEN.

XX (UYNT-) UNIV NIJMEGEN.

XX Linthorst HCM, Verpoorte R, Verberne MC, Moreno PRH;

PI Van Tegelen LJP, Willems GJ, Grose AF, Stuiver WD, Custers J;

PI Simons LH, Melchers LS, Bol JF;

XX WPI; 1999-610856/52.

DR P-PSDB; AAY50336, AAY50337, AAY50338, AAY50339.

XX Method for inducing pathogen resistance in plants.

XX Claim 5; Page 54-58; 66pp; English.

XX This invention describes a novel method for the induction of pathogen
CC resistance in plants, by transformation with an expression cassette
CC harboring a gene coding for an isochorismate synthase (ICS). A pathogen
CC inducible promoter can be used to drive expression of a heterologous
CC protein. The heterologous protein used in the method of the invention is
CC an antipathogenic protein e.g. chitinase, glucanase, osmotin, lectins,
CC saccharide oxidase, oxalate oxidase, magainins, toxins from *Bacillus*
CC *thuringiensis*, or antifungal proteins isolated from *Mirabilis jalapa*,
CC *Amaranthus*, *Raphanus*, *Brassica*, *Sinapis*, *Arabidopsis*, *Dahlia*, *Chicou*,
CC *Lathyrus*, *Clitoria*, *Allium* seeds, *Aralia* and *Impatiens* and albumin-type
CC proteins, such as thionin, napin, barley trypsin inhibitor, cereal
CC gliadin and wheat-alpha-amylase, or a protein that can induce a
CC hypersensitive response, such as Cf, Bs3 and Po proteins from tomato and
CC N-protein from tobacco. This sequence encodes the *Pseudomonas fluorescens*
CC isochorismate synthase orfA protein and the isochorismate pyruvate lyase
CC orfD protein which are described in the method of the invention. (Updated
CC on 27-AUG-2003 to correct OS field.)

XX Sequence 5057 BP; 1171 A; 1554 C; 1302 G; 1030 T; 0 U; 0 Other;

Query Match 2.6%; Score 46.6; DB 2; Length 5057;
Best Local Similarity 66.3%; Pred. No. 0.035;
Matches 67; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

Qy 1168 CATGTCGCCGACGTCTAGACGGCTCTGCGCTGCTCTGGAACAAAATGCCCAAGCGTTC 1227

Db 2570 CGTGCCCTTACCTTTTGGCCGGTACTACCAAGCCCGAACAAAATGCCCAAGCGTTC 2629

Qy 1228 GACGAGGAAGGCTATTACTGCTCCGGTGAATGCCATCAATT 1268

Db 2630 GATACGAGGGGCTACTACTCCGGCGACCTGTCAACT 2670

RESULT 12

ACA42560
ID ACA42560 standard; DNA; 1644 BP.

XX ACA42560;

XX 19-JUN-2003 (first entry)

XX Prokaryotic essential gene #24217.

XX Antisense; ds; prokaryotic essential gene; cell proliferation;

XX drug design; gene.

XX *Pseudomonas aeruginosa*.

XX WO200277183-A2.

XX 03-OCT-2002.

XX 21-MAR-2002; 2002MO-US009107.

XX 21-MAR-2001; 2001US-00815242.

XX 06-SEP-2001; 2001US-00948993.

XX 25-OCT-2001; 2001US-0342923P.

XX 08-FEB-2002; 2002US-00072851.

XX 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KU, Zyskind JW;

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX WPI; 2003-029926/02.

DR P-PSDB; AB038650.

XX New antisense nucleic acids, useful for identifying proteins or screening

PT for homologous nucleic acids required for cellular proliferation to

PT isolate candidate molecules for rational drug discovery programs.

XX Claim 14; SEQ ID NO 30430; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-regulated gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene

CC Product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1644 BP; 256 A; 627 C; 529 G; 232 T; 0 U; 0 Other;
Query Match 2.6%; Score 46.4; DB 8; Length 1644;
Best Local Similarity 53.3%; Pred. No. 0.025;
Matches 98; Conservative 0; Mismatches 86; Indels 0; Gaps 0;
OY 1081 GTGTGATGCTGTTACATTGGGCTGCAGCGCTGCTGCAGGTCAAGCTTCTCG 1140
DB 1093 CGGCGCGCTGTCGCGGACGAGAGTGGGTGTCAGCGCGCGCGAGGTGCGC 1152
OY 1141 GTGGATGGAAATTGGAAGGCGCTTTCATGTCGCGACGTCATGAGCGGCTACTGCGCT 1200
DB 1153 CCGGCGAGGTGGCGAAGTGAACCGTGGCGGCTCTTACACATCCGGGCTACTACGCG 1212
OY 1201 GCTCTGTAACAAATGCGCAAGCGTTGACAGAGGAGGCTATTACTGCTCGGTGATGCC 1260
DB 1213 CTGGCTGAACACACAGCGCAAGGCTTTCAGCGCGGAGCGCTTCTACCGACCGGACCGG 1272
OY 1261 ATCA 1264
DB 1273 GTCA 1276
RESULT 13
ADA71938/c
ID ADA71938 standard; DNA; 2000 BP.
XX
XX ADA71938;
XX
XX 20-NOV-2003 (first entry)
XX
XX Rice gene, SEQ ID 5263.
XX
XX Plant; bacterial infection; fungal infection; viral infection; rice;
XX gene; ds.
XX
XX Oryza sativa.
XX
XX OS
XX PN WO2003000898-A1.
XX
XX PD 03-JAN-2003.
XX
XX PF 22-JUN-2001; 2001WO-IB001105.
XX
XX PR 22-JUN-2001; 2001WO-IB001105.
XX
XX PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
XX PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
XX Katagiri F, Ques S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX WPI; 2003-175290/17.
XX
XX Identifying at least one gene involved in plant resistance or response to
XX pathogenic infection for conferring resistance or tolerance to a plant to
XX bacterial, fungal or viral infection by determining or detecting plant
XX gene expression.
XX
XX Claim 27; SEQ ID NO 5263; 899pp; English.

XX
CC The present invention relates to a method (M1) for identifying genes
CC involved in plant resistance or response to pathogenic infection. M1
CC comprises identifying a gene whose expression is significantly altered in
CC the incompatible interaction of plant gene expression relative to
CC expression of the gene in an uninfected plant. In a mutant plant that
CC does not express a gene associated with response to pathogenic infection,
CC or in a corresponding incompatible or compatible interaction. (M1) is
CC useful for conferring resistance to resistance or tolerance to a plant to
CC bacterial, fungal or viral infection. The present sequence was used to
CC illustrate the invention.
XX
SQ Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;
Query Match 2.6%; Score 45.6; DB 8; Length 2000;
Best Local Similarity 9.1%; Pred. No. 0.046;
Matches 65; Conservative 319; Mismatches 333; Indels 1; Gaps 1;
OY 659 TTTTGTGAAGACCGCGGCTGCTGCTGAGTGTGCGGTGGAACACACTTCCGCG 718
DB 720 KTKYASABWTKAKSKSYRRRRYRKKGWTRRYRWRSCMTRAMSKRKAKGASMK 661
OY 719 GCAACCAACATCGGATGCTGTGTATCAACGCGGACAGTACTTGAACAGGTA 778
DB 660 SCWYWRGASRMSYKSKYSCAGCKKTRWMTSYSTGMYSYKSMWTSMTSMYMGK 601
OY 779 AACCAACCGCCCAAGGTTGCGGACGCTTGCCAACTTGAGGAATCTTCCCACTG 838
DB 600 MTCMTYMSKGRSRSRSGMSRMYRMWKKRKRMYRMYKCTRRRCYMGWY 541
OY 839 CGTACCTCAGTGGCGGAAGGCTGGAGGAATTAAGGGTGCCTTGAGGAGACAGTA 898
DB 540 TMTTSSRMWTRRYARTRYSKRMYRYKRYKYVYVYGMWYKSYMYRYGYCKAK 481
OY 899 CCTGCGCAACGCTCTTCTGCTGCTGATGAAGCTTCTTCTGCGCGGCTGGTGT 958
DB 480 CCYAMCWKAAYSGMWYMYRKYSKMMSSTKYMMWYKCKRSMYKAGCYGCKMYTC 421
OY 959 CGCAAGGATCTGGATGCTTTGAGCGGCTCGTGAACAGCACTGTGTGAGGCATTC 1018
DB 420 SYGYMKWYTMGSYKSRCKYRMWYMYKGMWYMYSAVSMWYVYAKWYK 361
OY 1019 GCATGATGCGGCGGCTGGATGATGCGGACGAGCTGCTCTCTGCACTTTTACACCGGAC 1078
DB 360 RGTWMSYKSKYKXKCTWCMYKCMRYRMYKMKTKYSKRYCATYWCCTCYRK 301
OY 1079 CGCTGATGATGCTGTATGATGCTGCGGACGCTGCTGAGAGTCAAGCTCTGTC 1138
DB 300 RGMYSRSMWRTAGKWMRSMRCSRYSYKWKYKMKSYVMS-YGMARSSGTWBSA 242
OY 1139 CGGTGATGGAATTTGAAGGCGGCTTCCATGTCGCGACGTCATATAGCGCTACTGCGC 1198
DB 241 AKRTYKGYSTRBRKMMWRACRMYRACRYSRTSYCCSYGSGSKWYKMSGSMRTCS 182
OY 1199 GTGCTCTGAACAAATGCGGCTTGAACGAGGCTGATTTATGCTCCGCTGATG 1258
DB 181 SMCSCCTTCYGAACWSCMSWMTMSGCTTGMKRSKYSMCKKRYGCTCYGTYG 122
OY 1259 CCATCAATTTGCGAGATCTGCGGATCTCGAAGAGCTGTGATTTGACGCTGAATTG 1318
DB 121 YVRYCKMYKYSYKCYCYCWYWSYVRMYKMCNSCSGSMWSCAVSTSTSRMWSMY 62
OY 1319 CTGAAGACTTCAAGCTCTCTCAGGGGATTTGTCAAGCTTTGGCCATTGCGCACGGG 1376
DB 61 YAAWGMWCGSSGMWRMSKSKMYSKYSCTYTSKCTKRYVYCYVSSGSMWCTSG 4
RESULT 14
AAF26319
ID AAF26319 standard; DNA; 5451 BP.
XX
XX AAF26319;
XX

This Page Blank (uspto)

Query Match 100.0%; Score 1770; DB 6; Length 1770;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1770; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ATGCGTTCTCGAGGCGCTTTCTTCCCTTCCGGGTGGAATTCTTGAGCGCTTCGAGCAT 60
DB 1 ATGCGTTCTCTCAGAGGCGCTTCTTCCCTTCCGGGTGGAATTCTTGAGCGCTTCGAGCAT 60
QY 61 TGGGCTTAGACCCGTCAGAAACAACCTGCGTTGCTCGAGGGCGGCAATGGGGAATGG 120
DB 61 TGGGCTTAGACCCGTCAGAAACAACCTGCGTTGCTCGAGGGCGGCAATGGGGAATGG 120
QY 121 CGTGTATCAGTACCGGGAATGTTCCACACGTCGCGCATTCGACAGACCTTGCTT 180
DB 121 CGTGTATCAGTACCGGGAATGTTCCACACGTCGCGCATTCGACAGACCTTGCTT 180
QY 181 CCTTACGGAATATCGGAGAGCGTCCGCTTATCGTCTCTGGAATGACCTTGGAACAT 240
DB 181 CCTTACGGAATATCGGAGAGCGTCCGCTTATCGTCTCTGGAATGACCTTGGAACAT 240
QY 241 CTTACGCTGSCATTTGGGGCTATGATATGCGGCAATTCCTATTGCGCGGTCTCTGCT 300
DB 241 CTTACGCTGSCATTTGGGGCTATGATATGCGGCAATTCCTATTGCGCGGTCTCTGCT 300
QY 301 TATTCACGCTGTGCGAAAGATTTGGCGAGCTGCGTCAATCGTAGCTCTTTCGCAACG 360
DB 301 TATTCACGCTGTGCGAAAGATTTGGCGAGCTGCGTCAATCGTAGCTCTTTCGCAACG 360
QY 361 GGAAGTGTCTTTTCTGCGGATGAGCACCCTTTCAGGCGGCATTTGAGACATTTTCGCG 420
DB 361 GGAAGTGTCTTTTCTGCGGATGAGCACCCTTTCAGGCGGCATTTGAGACATTTTCGCG 420
QY 421 GAGAGCGTCCGCGCAATCTTCACTCGAGCGGAATTTGGCGGCGCGCAAGTGAATTT 480
DB 421 GAGAGCGTCCGCGCAATCTTCACTCGAGCGGAATTTGGCGGCGCGCAAGTGAATTT 480
QY 481 GACAGCTGCTGAGAGCAGCTGTGTGGATTGAGGCGAGATTAATGCTTTTGGCGCACTGCG 540
DB 481 GACAGCTGCTGAGAGCAGCTGTGTGGATTGAGGCGAGATTAATGCTTTTGGCGCACTGCG 540
QY 541 CCGGATACGATTTGCCAAGTCTTGTTCATCTTGGCTCTACAACTGCTTAAGCGGTG 600
DB 541 CCGGATACGATTTGCCAAGTCTTGTTCATCTTGGCTCTACAACTGCTTAAGCGGTG 600
QY 601 CCGACTACTCAGGGAATGCTCTGCGCAATCAGCAAGATGCTTCTGCAAACTTTCCGCTT 660
DB 601 CCGACTACTCAGGGAATGCTCTGCGCAATCAGCAAGATGCTTCTGCAAACTTTCCGCTT 660
QY 661 TTTGTGAGAGCGCCGCGGTGTGTGTGAGCTGTGTGCGTGAACCAACACTTTCGCGGCG 720
DB 661 TTTGTGAGAGCGCCGCGGTGTGTGTGAGCTGTGTGCGTGAACCAACACTTTCGCGGCG 720
QY 721 AGGCACAACATCGGCATCGTGTGTGAACAAGCGGCGCACTGTAACCTTGAACGAGTAA 780
DB 721 AGGCACAACATCGGCATCGTGTGTGAACAAGCGGCGCACTGTAACCTTGAACGAGTAA 780
QY 781 CCAACCGCCCAAGGGTTCGCGAGAGCTTTCGCAACTTGAAGCAAACTTCTCCCACTGCG 840
DB 781 CCAACCGCCCAAGGGTTCGCGAGAGCTTTCGCAACTTGAAGCAAACTTCTCCCACTGCG 840
QY 841 TACCTCACTGTGCGGAAGGCTGAGAGGAATTAAGTGGGTCCCTTGAAGCAGACAGTACC 900
DB 841 TACCTCACTGTGCGGAAGGCTGAGAGGAATTAAGTGGGTCCCTTGAAGCAGACAGTACC 900
QY 901 CTGCGGGAACGCTTCTTCCGCTGCAAGACGCTTCTTCCGCGGCGGTGGGTGTGCG 960
DB 901 CTGCGGGAACGCTTCTTCCGCTGCAAGACGCTTCTTCCGCGGCGGTGGGTGTGCG 960
QY 961 CAAGGATCTGGAGTCTTTTGAACCGGCTGCTGAACAGACAGTGTGGTGAAGCCATTGCG 1020
DB 961 CAAGGATCTGGAGTCTTTTGAACCGGCTGCTGAACAGACAGTGTGGTGAAGCCATTGCG 1020
QY 1021 ATGATGCGGGGTCTGGGCAATGACGGAAGTGTCTCTTCTGCACTTTTACACCGGACCG 1080
  
```

```

DB 1021 ATGATGCGGGGTCTGGGCAATGACGAGACGTCTTCTGCACTTTTACACCGGACCG 1080
QY 1081 CTGTGATGCGGTGCTTAACTTGGGCTGCGACGCGCTGCGAGGTCGAAGCTGTTCGCG 1140
DB 1081 CTGTGATGCGGTGCTTAACTTGGGCTGCGACGCGCTGCGAGGTCGAAGCTGTTCGCG 1140
QY 1141 GTGATGCGGAATTTGAGAGGCGCTTTCATGTGCTCCAGCTCATGAGCGGCTACTGCGGT 1200
DB 1141 GTGATGCGGAATTTGAGAGGCGCTTTCATGTGCTCCAGCTCATGAGCGGCTACTGCGGT 1200
QY 1201 GCTCTGAAACAATGCCCAAGCTTTCAGAGAGGAGGCTATTACTGCTCCGCTGATGCC 1260
DB 1201 GCTCTGAAACAATGCCCAAGCTTTCAGAGAGGAGGCTATTACTGCTCCGCTGATGCC 1260
QY 1261 ATCAAAATGCGAGATCTGCGCATCTCTCAAGAAAGGCTGATGTTTGAACGGTCAATTGCT 1320
DB 1261 ATCAAAATGCGAGATCTGCGCATCTCTCAAGAAAGGCTGATGTTTGAACGGTCAATTGCT 1320
QY 1321 GAAGCTTCAAGCTGTCTCAGGGGATATTTGTTCAGGCTTGGGCAATTCGCAAGCGGCG 1380
DB 1321 GAAGCTTCAAGCTGTCTCAGGGGATATTTGTTCAGGCTTGGGCAATTCGCAAGCGGCG 1380
QY 1381 GTTCTGAAAGGCGGCTCTTAACTGCTTGAACGTAAGTGTGCTGCTCTGATGCTGAATGC 1440
DB 1381 GTTCTGAAAGGCGGCTCTTAACTGCTTGAACGTAAGTGTGCTGCTCTGATGCTGAATGC 1440
QY 1441 CTTGATGCTGCTGTGTTCCGCGCTTCTGACCTGCGGCTTGTGCGGGCTGAGAAA 1500
DB 1441 CTTGATGCTGCTGTGTTCCGCGCTTCTGACCTGCGGCTTGTGCGGGCTGAGAAA 1500
QY 1501 GAGCGCTGCGAGCGCCGAGGTGCTTTCAGTGAAGCCGGTTCGGGCTGTGTTGCTGACTGG 1560
DB 1501 GAGCGCTGCGAGCGCCGAGGTGCTTTCAGTGAAGCCGGTTCGGGCTGTGTTGCTGACTGG 1560
QY 1561 CTCAAACTCACTCAATGAGAGCAATGCGCAATGCGATTCGATTAATGTGGTGAAGGCTC 1620
DB 1561 CTCAAACTCACTCAATGAGAGCAATGCGCAATGCGATTCGATTAATGTGGTGAAGGCTC 1620
QY 1621 CTCATACGCGCGCGCTGCAATTTGATTAAGGCGAGGTCACTGACAAAGGCTCATCAACAG 1680
DB 1621 CTCATACGCGCGCGCTGCAATTTGATTAAGGCGAGGTCACTGACAAAGGCTCATCAACAG 1680
QY 1681 CCGGCTGTTTTCGAATGAGCGGTGCGGGAAGTTGATGCGCTGTATGCTGTGTAAGATCAA 1740
DB 1681 CCGGCTGTTTTCGAATGAGCGGTGCGGGAAGTTGATGCGCTGTATGCTGTGTAAGATCAA 1740
QY 1741 TCCATGCTGCTGACGAGGCGCACACTGTGA 1770
DB 1741 TCCATGCTGCTGACGAGGCGCACACTGTGA 1770

```

RESULT 2
 PSP238746 4166 bp DNA linear BCT 05-NOV-1999
 LOCUS
 DEFINITION
 Pseudomonas sp. vdh (partial), fcs, aat and mac (partial) genes, strain HR199.
 ACCESSION
 AJ238746
 VERSION
 AJ238746.1 GI:6273620
 KEYWORDS
 aat gene; beta-ketothiolase; fcs gene; feruloyl-CoA synthetase; mac gene; methyl-accepting chemotaxis protein; vanillin dehydrogenase; vdh gene.
 SOURCE
 Pseudomonas sp.
 ORGANISM
 Pseudomonas sp.
 Bacteria; Proteobacteria.
 REFERENCE
 1 Overhage, J., Priefer, H. and Steinbuechel, A. Biochemical and genetic analyses of ferulic acid catabolism in Pseudomonas sp. Strain HR199 Appl. Environ. Microbiol. 65 (11), 4837-4847 (1999)
 JOURNAL MEDLINE
 PUBMED
 10543794
 2 (bases 1 to 4166)
 REFERENCE

AUTHORS
TITLE
JOURNAL

Priefert, H.
Submitted (04-MAY-1999) Priefert H., Westfälische
Wilhelms-Universität Münster, Institut fuer Mikrobiologie,
Corrensstrasse 3, Münster, D-48149, GERMANY

FEATURES
SOURCE

Location/Qualifiers

1..4166

/organism="Pseudomonas sp."

/mol_type="genomic DNA"

/strain="HK199"

/db_xref="taxon:306"

/country="Indonesia"

1..170

/gene="vdh"

<1..128

/function="NAD dependent oxidation of vanillin to vanillic
acid"

/codon_start=3

/evidence=experimental

/transl_table=11

/product="vanillin dehydrogenase"

/protein_id="CAB60225.1"

/db_xref="GI:6273621"

/db_xref="GOA:O9RLD0"

/translation="QMFGVKSSTGSGFSRASIEHFTQLRMITONGRRHPI"

138..170

/gene="vdh"

/function="putative terminator"

join(373..378,388..2157)

/gene="fcsa"

373..378

/gene="fcsa"

388..2157

/gene="fcsa"

/function="ATP dependent conversion of ferulic acid to
feruloyl-Coenzyme A"

/codon_start=1

/evidence=experimental

/transl_table=11

/product="feruloyl-CoA synthetase"

/protein_id="CAB60226.1"

/db_xref="GI:6273622"

/db_xref="GOA:O9RLD9"

/db_xref="TrEMBL:O9RLD9"

/translation="MRSLEALPPGRILRLHEMAKTRBEQTCVAAANGEMRRIS
YAEFNVRAIAOSILPYGLSAERPLIVSGNDLEHLQAFAMVAGIPLYSPAYS
LSODLAKLTHIVGLQPLVFPADAPFORAETILPDVPAIFETGELAGRTVAF
DLLEOFGIEADAPAPATGPTIAKELFTSGSTKLPKAVPTTORMCANOMLOT
PVGEPEPVLDMLPMHTTGGSHNIGIVLNGSTYLLDGGKPTAGCPAETRLSEI
SPFAYLVTPKGEMLVGLERDSTLERFPAKMLFPALAGLSQGIWDRIDRAEHC
GGRIRMAAGIEMTERPSCFTTGPISMGYIGLPAPGCEVKLVPDGKLEGFHBP
HMGVIRMAEONQAFDEBEGYCSGDAIGLADPADPDKGIMFGRLAEDEKSSGVF
VSSPRLTRAVLEGGSYLVLVVNAAPRECIIGLVFPRLLDCRALSGIGBASAEVL
ASPPVRAWFPADMLKRLREATGNASRLIMWGLDTPPSIDKGEVTDGSGINQRAVLQW
RSKAKVDLTYGGEOSMLRDEATL"

2143..2146

/gene="aat"

2143..2146

/gene="aat"

2154..3449

/gene="aat"

2154..3449

/gene="aat"

/codon_start=1

/transl_table=11

/product="putative beta-ketothiolase"

/protein_id="CAB60227.1"

/db_xref="GI:6273623"

/db_xref="TrEMBL:O9RLD8"

/translation="MSWGGAYSAFSDTALVAARTPMIDCGALSLVSPIDLGKVA
REVLMRASLEPQWVDSVLASGMAQAFDAYLLPRHIGLYSGVPSVPAIVGRICTG

gene
RBS
CDS

3521..4166
3521..3525
/gene="mac"
3535..4166
/gene="mac"
/codon_start=1
/transl_table=11
/product="putative methyl-accepting chemotaxis protein"
/protein_id="CAB60228.1"
/db_xref="GI:6273624"

ORIGIN

Query Match 100.0%; Score 1770; DB 1; Length 4166;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1770; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

FEILRQGEIOISQADHVLCAVAESMRNPJASYTHRGFRGLGAPVEFKDLMEALPD
PARGIDMTATAEMLARLYGTRGEEANVAVSSEPRALRABQEKWIDOEIYAVDEQD
LEGNSRAITLPRKAKILYTVIRGLAVFPAISLURKRVHSGVOTANSCAVNDGAA
ALVARESSATQPVILATLAVSIVGIEBEHGLGAPRILRLARSLDIDLEFIN
EQAQAQVLAQVHIEHSKINIWGGAIALGHPILAAAGLRLCMTLAIQLOANNRYGI
ASACIGGGQGMALVLEBNPHGSSSARSMSINRVHPLDS"
3521..4166
/gene="mac"
3521..3525
/gene="mac"
3535..4166
/gene="mac"
/codon_start=1
/transl_table=11
/product="putative methyl-accepting chemotaxis protein"
/protein_id="CAB60228.1"
/db_xref="GI:6273624"

/translation="MISFARMASLGVQAKLALFALVLCVGLIVTGTFYSVHTLSG
LVKASAIAGELRAKIOELKVLBORALFIDEGSKORSILISQVIAVNDIAIDIPQ
RGSSELKPAASREASYSIEVSNNAAIDLGSEBPDALMDVADKLVNEYEQLSAV
NALMGLIEDQNEKRVPLIYMLGVTILFTLMSAYSVFISRLQVPLKST"

1 ATGGCTTCTCTCGAGGCGCTTCTTCCCTCCGGGTGAAATCTTGAGCGTTCGACAT 60
388 ATGGCTTCTCTCGAGGCGCTTCTTCCCTCCGGGTGAAATCTTGAGCGTTCGACAT 447
61 TGGGCTTAAGACCCGTCACCAACAACCTGGTTCCTCCAGGCGGCAATGGGAATGG 120
448 TGGGCTTAAGACCCGTCACCAACAACCTGGTTCCTCCAGGCGGCAATGGGAATGG 507
121 CGTGTATCAGCTACGCGGGAATGTTCCACAACGTCGCGCCATCGACAGAGCTTCTT 180
508 CGTGTATCAGCTACGCGGGAATGTTCCACAACGTCGCGCCATCGACAGAGCTTCTT 567
181 CTTACGGAATATCGGAGAGCGTCGCTTATCGTCTCTGGAATGACCTGGAACAT 240
568 CTTACGGAATATCGGAGAGCGTCGCTTATCGTCTCTGGAATGACCTGGAACAT 627
241 CTTACGGAATATCGGAGAGCGTCGCTTATCGTCTCTGGAATGACCTGGAACAT 300
628 CTTACGGAATATCGGAGAGCGTCGCTTATCGTCTCTGGAATGACCTGGAACAT 367
301 TATTCACTGCTGTCGCAAGATTGGCGAAGCTGCTGCAATCGTAGGCTTCTGCAACG 360
688 TATTCACTGCTGTCGCAAGATTGGCGAAGCTGCTGCAATCGTAGGCTTCTGCAACG 747
361 GGAATGCTCTTGTGTCGCAATGTCGCAACCTTTCAGGCGCAATGAGCAATTCGCG 420
748 GGAATGCTCTTGTGTCGCAATGTCGCAACCTTTCAGGCGCAATGAGCAATTCGCG 807
421 GAGCACTGTCGCGCAATCTTCACTCGAGCGGAATGGCGGCGGCGCAAGTGAATTT 480
808 GAGCACTGTCGCGCAATCTTCACTCGAGCGGAATGGCGGCGGCGCAAGTGAATTT 867
481 GAGCACTGTCGCGCAATCTTCACTCGAGCGGAATGGCGGCGGCGCAAGTGAATTT 927
868 GAGCACTGTCGCGCAATCTTCACTCGAGCGGAATGGCGGCGGCGCAAGTGAATTT 1047
601 CCGATACGATTCGCAAGTCTTGTGCACTTTCGCTTCAACAATGCTTAAAGCGGTG 600
928 CCGATACGATTCGCAAGTCTTGTGCACTTTCGCTTCAACAATGCTTAAAGCGGTG 987
601 CCGATACGATTCGCAAGTCTTGTGCACTTTCGCTTCAACAATGCTTAAAGCGGTG 660
988 CCGATACGATTCGCAAGTCTTGTGCACTTTCGCTTCAACAATGCTTAAAGCGGTG 1047
661 TTGGTGAAGAGCCGCGGTGCTGCTGGAATGCTGCTGGAACCAACCTTCGCGCG 720

```
Db 1048 TTTGTAAGAGCGCGGTGCTGTGAGCTGGTTGCCGTGAACACACTTCGCGGCG
Qy 721 AGCCACAACATCGGCATCGTGTGTATCAAGCGCGGACAGTACCTTGACGACGGTAA 780
Db 1108 AGCCACAACATCGGCATCGTGTGTATCAAGCGCGGACAGTACCTTGACGACGGTAA 1167
Qy 781 CCAACGCCCAAGGGTTCGCGGAGAGCTTCGCACTTGAGCGAAATCTCTCCACTGCG 840
Db 1168 CCAACGCCCAAGGGTTCGCGGAGAGCTTCGCACTTGAGCGAAATCTCTCCACTGCG 1227
Qy 841 TACCTCACTGTGCGGAAAGCTGAGAAATTAAGTGGTCCCTTGAGCAGACAGTACC 900
Db 1228 TACCTCACTGTGCGGAAAGCTGAGAAATTAAGTGGTCCCTTGAGCAGACAGTACC 1287
Qy 901 CTGCGGGAAGCGCTTCCTGCTGCGATGAGACGTCTCTTCGCGGCGGTGCGGTGTCG 960
Db 1288 CTGCGGGAAGCGCTTCCTGCTGCGATGAGACGTCTCTTCGCGGCGGTGCGGTGTCG 1347
Qy 961 CAAGGATCTGGATCGTTTGAACCGGGTCTGAAACAGACAGTGTGAGCGCAATTGCG 1020
Db 1348 CAAGGATCTGGATCGTTTGAACCGGGTCTGAAACAGACAGTGTGAGCGCAATTGCG 1407
Qy 1021 ATGATGCGCGGTCTGGGATGACGAGAGCTGCTCTTCTGCACTTTTACCCGAGACG 1080
Db 1408 ATGATGCGCGGTCTGGGATGACGAGAGCTGCTCTTCTGCACTTTTACCCGAGACG 1467
Qy 1081 CTGTCGATGCGGTATACATTTGGGCGCGCGGCTGAGCGGCTGAGAGGTCAAGTCCG 1140
Db 1468 CTGTCGATGCGGTATACATTTGGGCGCGCGGCTGAGCGGCTGAGAGGTCAAGTCCG 1527
Qy 1141 GTGATGGAAGAAATTTGAAGGCGCTTTCATGATCCGACGTCATGAGCGGCTACTGCG 1200
Db 1528 GTGATGGAAGAAATTTGAAGGCGCTTTCATGATCCGACGTCATGAGCGGCTACTGCG 1587
Qy 1201 GCTCTGGAAGAAATTTGCGCAAGCGCTTTCATGAGAGAGGCTATTAAGTCTCGGTATGCC 1260
Db 1588 GCTCTGGAAGAAATTTGCGCAAGCGCTTTCATGAGAGAGGCTATTAAGTCTCGGTATGCC 1647
Qy 1261 ATCAAAATGGCAGATCCGCGATCCCTCAGAAAGGTGATGTTTGAAGCGTTCGAAATTC 1320
Db 1648 ATCAAAATGGCAGATCCGCGATCCCTCAGAAAGGTGATGTTTGAAGCGTTCGAAATTC 1707
Qy 1321 GAAGACTTCAAGCTGCTCAGAGGGTATTTGTCAAGCGTTGAGGCAATTCGCAAGCGGCG 1380
Db 1708 GAAGACTTCAAGCTGCTCAGAGGGTATTTGTCAAGCGTTGAGGCAATTCGCAAGCGGCG 1767
Qy 1381 GTTCTGGAAGGCGGCTTTCATGCTGAGAGCTGAGTGTGCTGCTCTGATCGTGAATGC 1440
Db 1768 GTTCTGGAAGGCGGCTTTCATGCTGAGAGCTGAGTGTGCTGCTCTGATCGTGAATGC 1827
Qy 1441 CTGGAATGCTCGTGTTCGCGGCTTTCGACGTCGCGGCTTTCGAGGCTGAGGAA 1500
Db 1828 CTGGAATGCTCGTGTTCGCGGCTTTCGACGTCGCGGCTTTCGAGGCTGAGGAA 1887
Qy 1501 GAGGCGTCGAGCGCGAGAGTCTTTCAGAGTGAAGCGGCTTTCGAGGCTGAGGAA 1560
Db 1888 GAGGCGTCGAGCGCGAGAGTCTTTCAGAGTGAAGCGGCTTTCGAGGCTGAGGAA 1947
Qy 1561 CTGGAAGCTCAATGAGAGAGCACTGAGAGTCCAGTGTGATCATGTGTGTGAGGCTC 1620
Db 1948 CTGGAAGCTCAATGAGAGAGCACTGAGAGTCCAGTGTGATCATGTGTGTGAGGCTC 2007
Qy 1621 CTGGAATGCGCGCGTGTGATTAAGAGGAGAGTCACTGACAGAGGCTGATCAACAG 1680
Db 2008 CTGGAATGCGCGCGTGTGATTAAGAGGAGAGTCACTGACAGAGGCTGATCAACAG 2067
Qy 1681 CGGCGTGTGTAATGAGCGGTGCGGAAAGTGTGATGCGTGTATCGTGTGAAATCAA 1740
Db 2068 CGGCGTGTGTAATGAGCGGTGCGGAAAGTGTGATGCGTGTATCGTGTGAAATCAA 2127
Qy 1741 TCCATGCTGTGTAACGAGGCCACACTGTGA 1770
```

```
Db 2128 TCCATGCTGTGTAACGAGGCCACACTGTGA 2157
RESULT 3
A92090
LOCUS A92090 32679 bp DNA linear PAT 22-JAN-2000
DEFINITION Sequence 1 from Patent EP0845532.
ACCESSION A92090
VERSION A92090.1 GI:6740886
KEYWORDS
SOURCE Pseudomonas sp.
ORGANISM Pseudomonas sp.
REFERENCE Bacteria; Proteobacteria.
AUTHORS 1 (bases 1 to 32679)
Priefert, H.D. and Rabenhorst, J.D.
TITLE Enzymes for the synthesis of coniferyl alcohol, coniferyl aldehyde,
JOURNAL ferulic acid, vanillin, vanillic acid and their applications
HARMANN & REIMER GMBH (DE)
FEATURES
source
1..32679
/organism="Pseudomonas sp."
/mol_type="unassigned DNA"
/strain="HR19"
/db_xref="taxon:306"
3146..4000
/gene="ORF1"
3146..4000
/gene="ORF1"
/codon_start=1
/translation="MIAITGASGOLRLTALRLPASEIILAVRPDNNKAGDLTAR
/db_xref="GI:6740887"
/translation="MIAITGASGOLRLTALRLPASEIILAVRPDNNKAGDLTAR
SLHADKSLGLATEHRTDQLTSGIPIHVLIRNMTHTNTAGIPVALVHGVLLGC
ADDDGLASARADYAEAAAVLVTEGQAGRVLEGEPAVYITLAEVAVPOGKTIV
YENLSQSDYRSALISAGLPDFAALLADSDAAMAKYLPDSSGDSRLIGRPPTPMS
AIAAATGR"
ORIGIN
Query Match 100.0%; Score 1770; DB 6; Length 32679;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1770; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATGCGTCTCTCGAGGCGCTTCCCTCCCGGCTGCAATTCCTGAGGCTCGAGCAT 60
Db 2128 ATGCGTCTCTCGAGGCGCTTCCCTCCCGGCTGCAATTCCTGAGGCTCGAGCAT 21347
Qy 61 TGGGCTAAGACCCGTCAGAAACAACTGCTGTCGCAAGGCGGCAATGGGAATGG 120
Db 21348 TGGGCTAAGACCCGTCAGAAACAACTGCTGTCGCAAGGCGGCAATGGGAATGG 21407
Qy 121 CCGTATCAAGCTACCGGAAATGTTCCCAACAGTCCGCGCATCGCACAGAGCTTGC 180
Db 21408 CCGTATCAAGCTACCGGAAATGTTCCCAACAGTCCGCGCATCGCACAGAGCTTGC 21467
Qy 181 CCTTACGACATACGAGAGCGTCCGCTGTTATCGTCTGGAATGACCTGGAACAT 240
Db 21468 CCTTACGACATACGAGAGCGTCCGCTGTTATCGTCTGGAATGACCTGGAACAT 21527
Qy 241 CTTACAGTGGCAATTTGGGGCTATGTAAGCGGCAATTCCTATTGCCGCTGCTCTG 300
Db 21528 CTTACAGTGGCAATTTGGGGCTATGTAAGCGGCAATTCCTATTGCCGCTGCTCTG 21587
Qy 301 TATTCACTGCTGTGCGAAGATTTGGAGAGCTGCGTCAATCGTAGTCTTCTGCAACG 360
Db 21588 TATTCACTGCTGTGCGAAGATTTGGAGAGCTGCGTCAATCGTAGTCTTCTGCAACG 21647
Qy 361 GAGCTGCTTTGCTGCGATGAGCACTTTCAGCGGCAATTGAGACCATTTGCGCG 420
Db 21648 GAGCTGCTTTGCTGCGATGAGCACTTTCAGCGGCAATTGAGACCATTTGCGCG 21707
```

QY 421 GAGCAGTCGCCGCAATCTTCACTCGAGGCGAATTGGCCGGCGCGGCGACGGTGATTTT 480
 DB 21708 GAGCAGTCGCCGCAATCTTCACTCGAGGCGAATTGGCCGGCGCGGCGACGGTGATTTT 21767
 QY 481 GAGCAGTCGTGGAGCAGCCGTGTGGATTGAGGCAATATATGCTTTGCGGCAACTGGC 540
 DB 21768 GAGCAGTCGTGGAGCAGCCGTGTGGATTGAGGCAATATATGCTTTGCGGCAACTGGC 21827
 QY 541 CCGGATACGATTGGCAAGTTCTTGTTCATCTTGTGCTTACCAAACTGCTTAAGCCGTG 600
 DB 21828 CCGGATACGATTGGCAAGTTCTTGTTCATCTTGTGCTTACCAAACTGCTTAAGCCGTG 21887
 QY 601 CCGGATACGATTGGCAAGTTCTTGTTCATCTTGTGCTTACCAAACTGCTTAAGCCGTG 660
 DB 21888 CCGGATACGATTGGCAAGTTCTTGTTCATCTTGTGCTTACCAAACTGCTTAAGCCGTG 21947
 QY 661 TTGTGTAGAGCCGCGGT 720
 DB 21948 TTGTGTAGAGCCGCGGT 22007
 QY 721 AGCCACAACATCGGATCTGT 780
 DB 22008 AGCCACAACATCGGATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 22067
 QY 781 CCAACGCCCAAGGGGTGCGCGAGAGCGCTTGGCAACTTGAAGCGAATCTCTCCCACTGGC 840
 DB 22068 CCAACGCCCAAGGGGTGCGCGAGAGCGCTTGGCAACTTGAAGCGAATCTCTCCCACTGGC 22127
 QY 841 TACCTCACTGTGCGAAAGGCTGGAGGAAATTAGTGGGTGCTTGAAGCGAGACATACC 900
 DB 22128 TACCTCACTGTGCGAAAGGCTGGAGGAAATTAGTGGGTGCTTGAAGCGAGACATACC 22187
 QY 901 CTGCGCGAAGCGCTTCTTCTGCTCGCATAGAGTGTCTTCTGCGCGCGCTGGTGTGCG 960
 DB 22188 CTGCGCGAAGCGCTTCTTCTGCTCGCATAGAGTGTCTTCTGCGCGCGCTGGTGTGCG 22247
 QY 961 CAAGGAGCTGGGATCGTTTGGACCGGGTGGCTAAGAGACATGTGGTGAAGCGCATTCGC 1020
 DB 22248 CAAGGAGCTGGGATCGTTTGGACCGGGTGGCTAAGAGACATGTGGTGAAGCGCATTCGC 22307
 QY 1021 ATGATGCGGGGTCTGGGATGACGAGACATGCTCTTCTGCACTTTTACCAACCGGACCG 1080
 DB 22308 ATGATGCGGGGTCTGGGATGACGAGACATGCTCTTCTGCACTTTTACCAACCGGACCG 22367
 QY 1081 CTGTGATGCTGTGTATCATTTGGGCTGCCAGCGCTGGCTGGAGGTCAAGCTGTTCG 1140
 DB 22368 CTGTGATGCTGTGTATCATTTGGGCTGCCAGCGCTGGCTGGAGGTCAAGCTGTTCG 22427
 QY 1141 GTGATGCGGAAATTGAAAGGCGTTTCCATGCTCGGACGTCATGAGCGGCTACTGGCGT 1200
 DB 22428 GTGATGCGGAAATTGAAAGGCGTTTCCATGCTCGGACGTCATGAGCGGCTACTGGCGT 22487
 QY 1201 GCTCTGMAAATAATGCGCAAGGCTTGCAGAGAGAGGCTATTACTGCTCGGTGATGCC 1260
 DB 22488 GCTCTGMAAATAATGCGCAAGGCTTGCAGAGAGAGGCTATTACTGCTCGGTGATGCC 22547
 QY 1261 ATCAAAATGGCAATCTCTCGCATCTCAAGAAAGTGTGATGTTGAAGCTCGAATGCT 1320
 DB 22548 ATCAAAATGGCAATCTCTCGCATCTCAAGAAAGTGTGATGTTGAAGCTCGAATGCT 22607
 QY 1321 GAAGACTTGAAGCTCTCAAGGCGTATTGTGACGCTTGGCGCATGCGCAACGCGGCGG 1380
 DB 22608 GAAGACTTGAAGCTCTCAAGGCGTATTGTGACGCTTGGCGCATGCGCAACGCGGCGG 22667
 QY 1381 GTTCTGMAAGCGGCTTCAAGCTCTGAGAGATGATGTTGCTGCTGCTGATTCGTAATGC 1440
 DB 22668 GTTCTGMAAGCGGCTTCAAGCTCTGAGAGATGATGTTGCTGCTGCTGATTCGTAATGC 22727
 QY 1441 GTTGAATGCTCTGTTTCCGCGCTTCTCTGACTGCGCGCTTGTGTGGGGCTAGAGAAA 1500
 DB 22728 GTTGAATGCTCTGTTTCCGCGCTTCTCTGACTGCGCGCTTGTGTGGGGCTAGAGAAA 22787
 QY 1501 GAGGCGTCGAGCGGAGGTGCTTGCAGATGAGCGCGTGTGGGCTGTGCTGACTGG 1560

DB 22788 GAGGCGTCGAGCGCGAGAGTGTCTTGCAGTGAAGCCGGTGTGGGCTGTGCTGACTGG 22847
 QY 1561 CTCAACGACTCAATTCAGAGAGCACTGGCAATGCCATGTCATCATGTGGTGAAGGCTC 1620
 DB 22848 CTCAACGACTCAATTCAGAGAGCACTGGCAATGCCATGTCATCATGTGGTGAAGGCTC 22907
 QY 1621 CTGATTCGCGCGCTGTGATTTGATTAAGGCGAGGTCTCAACAAGGCGCTGATCAACAG 1680
 DB 22908 CTGATTCGCGCGCTGTGATTTGATTAAGGCGAGGTCTCAACAAGGCGCTGATCAACAG 22967
 QY 1681 CGCGCTTTTGGCAATGCGCGTGGCGCAAGTTGATCGCTGATCGTGTGAAGATCAA 1740
 DB 22968 CGCGCTTTTGGCAATGCGCGTGGCGCAAGTTGATCGCTGATCGTGTGAAGATCAA 23027
 QY 1741 TCCATGCTGCGTGAAGAGGCGCACACTGTGA 1770
 DB 23028 TCCATGCTGCGTGAAGAGGCGCACACTGTGA 23057

RESULT 4
 PFL536325
 LOCUS
 DEFINITION
 Pseudomonas fluorescens ferr gene, ech gene, vdh gene, fcs gene, aat gene and fcd gene.
 ACCESSION
 VERSION
 A536325.1 GI:40789116
 KEYWORDS
 aat gene; beta-ketothiolase; ech gene; fcd gene; fcs gene; ferr gene; feruloyl-CoA dehydrogenase; feruloyl-CoA hydratase; feruloyl-CoA synthetase; transcriptional regulator; vanillin dehydrogenase; vdh gene.
 SOURCE
 ORGANISM
 Pseudomonas fluorescens
 Pseudomonas fluorescens
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 Ruzzi, M.
 Cloning and Characterization of Ferulate Catabolism Genes from Pseudomonas fluorescens Bfl3
 Unpublished
 2 (bases 1 to 9477)
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 Ruzzi, M.
 Direct Submission
 Submitted (09-JAN-2003) Ruzzi M., Agrobiology and Agrochemistry, DNBAC - University of Tuscia, via C. de Dellis, snc, Viterbo - 01100, ITALY
 FEATURES
 source
 1..9477
 /organism="Pseudomonas fluorescens"
 /mol_type="genomic DNA"
 /strain="Bfl3"
 /isolation_source="soil"
 /db_xref="taxon:294"
 /environmental_sample
 /country="Italy:Lazio"
 /complement(1297..1785)
 /gene="ferr"
 /complement(1297..1785)
 /gene="ferr"
 /codon_start=1
 /evidence=experimental
 /transl_table=11
 /product="transcriptional regulator"
 /protein_id="CAD60265.1"
 /db_xref="GI:40789117"
 /translation="MAKPAPLADPREAPANAEOVAPLSDBELIGYAMRAOLKLF
 ONLIGLSADHBPAPQFASALATIDSNPGMOWDLARALAIPEQVVPPLNKJESSRLA
 VNRCKPDKRSIGIFLSTKGTLLKELKQIAOSVDATSALESREBELRLKKYV
 QD"
 1994..2824
 /gene="ech"
 1994..2824
 /gene="ech"
 /gene="ech"

```
/codon_start=-1
/evidence=experimental
/transl_table=11
/product="feruloyl-CoA hydratase"
/protein_id="CAD60266.1"
/db_xref="GI:40788118"
/translation="MSGNGRWTTTVEIEDGIAMVILNPEKRNAMSPITLNEMIDV
LETLBODPAGVILVTGAEAMTWAGDVLKEFREVDAPEIIOEKIRREASQWMLL
RMARPTLIAMVNMKCFGGFSPVACDLAI CADEATFGSEIMWGI.PPGNLVSKAAD
TVHROSLEYIMTKFTGGOKAEMGVINDSVPLARLREYITELANLLEKNVIVRA
AKHGFRCRELTEWQEDVIYALUDSRILDTGREGKQKPLDDKS1KPGIQAYKR
"
2921. .4369
/gene="vch"
/2921. .4369
/gene="vch"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="vanillin dehydrogenase"
/protein_id="CAD60267.1"
/db_xref="GI:40788119"
/translation="MLDVLPLIGGOSRPARDGTLRRNPVNGAVSVYAAATLDDAD
AAYAAQAAPPAALAPNRRSRRLKAHQLOARSGEPTAAGETGAMAMTGVNVR
LAANMLREASMTQINGEVTPSDVGSFAMALRQCGVVLGIAPMVAIVITRAIA
MPLACNTVYLKASSELSPAYHRLIGVLDAGAGDVVNVISNAPVDAATVIRLTAN
PAVRVNTFGSTHVRIGVELARHLKPLBELGKAPLILDDADLDAVAALAGA
YFNOGOICMSTERLIYDAKVAADFVAKLAQVATLRAGDPSADSVIGSLVDVAGTR
IKALIDALSKGARLVVGGQIDGSIIOPTLIDGVTSMRLYREESFGPAVILRGSD
BELRLANDSEFGLSNAISRPDGRALALAQVRESGICHTNGPTVDEAMPPGVKS
SGTSGFGKASIHFTQLRWVTLQHGPRHPI"
4628. .6397
/gene="fcs"
/4628. .6397
/gene="fcs"
/codon_start=1
/transl_table=11
/product="feruloyl-CoA synthetase"
/protein_id="CAD60268.1"
/db_xref="GI:40788120"
/translation="MRSIKPLIALPRLILDRLVHMAVRPEQFIAREAGDMRLVS
YGMDSVRAIAGILRYGISADKPLALISGNDIEHLQALAMVAGTIPRCPSAYS
LISQDPKLHNVCDLQPGLVFVSDASATQRAIDAVLPETPLISVGGVPGKQSF
ASLILAPGKAADAAFAATGPDSIAKFLFTSGSTKLPKVITTTQRMICANQOMLQTF
PVFGEPPVYLVDMLPMNHTFGSGHNVIYLVNGSTKPLDGKPTAGFAETLKEI
SPFAYLTVPKGMEILYNALBODALRECFKRMSLFFPFAAGLSQWMDRLDVAQHS
CGRIITMMAGICTEAPASCTFTTGPLSMAGYGLPAPGCEVTLVVDGKLBGRFPG
HIMPGYRAAQCTAYFPDAOGFYCSGDALKADADPOLGMDGRIABDFKLSGVF
VSGPLNRPAVLEGAIPYVDLVVAAPDRECLALVPRLYECRLGLGAEAEQVL
ASAPVQWFGDWLQCLNRSASGASRLLEWIALQVEPASIDREITDKGSIINDRAVLQW
RAQVENVLYRGSPSILRAEPK"
6394. .7641
/gene="aat"
/6394. .7641
/gene="aat"
/codon_start=1
/transl_table=11
/product="putative beta-ketochiolase"
/protein_id="CAD60269.1"
/db_xref="GI:40788121"
/translation="MSGAGYSGFADVALIEAVRTWVDLGDIAEISPIDGIKVR
EVLARAGIPLQAVDSVLASMAQSFADALPHRIGLYSGVGQAVFALVQORLCATGF
BILRQABOILRGVQLALCVASRSMRNPIAATHRGGRGLGAQVFXDFLMALYDP
ALGLDMITADNLARRYGLTREAVDGATCSHORALAAQAFPAEIVASQOERL
DYGAGIRLPRGVERVSDSHPRSEBNDIARLTIHFGQVOTAGNSCAVVDGAAAL
LVGRASCSDRPLALIRASAVYGVABEPNGIGAPAIRILLQKMEBAEQIRSAEVI
NKAQSOVLAVAEILGDGRLVNVQGSIALGHPILASGLRLVTLARQRENNRYG
IAAACVGGQGMALILENPAMRA"
7701. .9455
/gene="fcd"
/7701. .9455
/gene="fcd"
/codon_start=1
```

```
/transl_table=11
/product="putative feruloyl-CoA dehydrogenase"
/protein_id="CAD60270.1"
/db_xref="GI:40788122"
/translation="MMSYQAPLRDNOFVLEHLEAPAMRRSPAFALDPLALQYLE
QARFSGVLAFLAMSGDRQGMADQGVSTREGRADATRAYTEGGWPAALCEALAG
QGLPOLDLAALDEMIYASNRHAMAMTGIAGAVLCLKTGAEPLOQRYVPAI1SGASL
PTMCLTEPLIASDVGLLRABEPADGSRVSGSKLFI SGGBDLTDNTLILVLAFLP
DABEGSRGISLFLVPRLEQGDNGVRCNGIEHKMG1KGSATCSLVFDAQGWLEA
NGLIAMFVMMNSARLHVLOGIGHVAAWONALAYAGROQWRALRPAQGAQOAZD
PIRYHPAMRVLLELRATSEGRALIGYMAAHLDDOENHPEAPROADLOALPI
IRAFTEQGFROASNALQVFGCYGYVAEPAIEOTLDSIAMIYBSNEI QANDILIR
KVLGDGGGFALLLEQRDEABQGAAGVACAGAGLADLAEALQGLVGVQWALSD
ABPYRAADDFLALCELISLALAMARAARVSRNLPGDPLRKAKTATAGFFSYLLPQ
VEQMTAVASQAALAFI"
ORIGIN
```

Query Match 55.8%; Score 987.6; DB 1; Length 9477;
Best Local Similarity 72.4%; Pred. No. 2,4e-226;
Matches 1281; Conservative 0; Mismatches 489; Indels 0; Gaps 0;

QY	1	ATGCGTTCCTGAGGCGCTTCCCTCCCGGGTCSAATTCTTGAGCGTCCGAGCAT	60
DB	4628	ATGCGCTCGTGAAGCCGCTTGCGGCTTGCCGATCGCTTGTCATGCTGCTGTCAC	4687
QY	61	TGGGCTTAAGACCCGTCAGAAACAACCTCGTTGCTGCCAGGGCGGCAATGGGAAATGG	120
DB	4688	TGGGCGCGGTGGCGCCCGAGAGACCTTCATTGCGCCCGCGAGGCGCGGTGGCGACTGG	4747
QY	121	CGTCGTATACGTACGCGGAAATGTTCCACAAAGTCGCGGCCATCGGCACAGACTTGCTT	180
DB	4748	CGCCTTGTCAGCTATATGGAGAGATGTCGACAGGTGCGGCGCATCGCCAGGGCTGCTT	4807
QY	181	CCTTACGGAATATGCGACAGCGCTCCGCTGCTTATCGTCTCTGGAATAGCACTGGAACAT	240
DB	4808	CGTTACGGCTGTGCGCCAGCAAGCCGCTGGCGCTCTCTCGGGCAATGACATCGAACAC	4867
QY	241	CTTCACTGCGCAATTTGGGGCTATGTAATGCGGGCAATTCCTATTTCCCGGTGTCCTGCT	300
DB	4868	CTGCAACTGGCCCTGGGGCGAGATGTACGCGGGATTCCTGATGCGCGGTGTCGGGCC	4927
QY	301	TATTCATCTGCTGTGCGAAGATTTGGGAGAGCTGCGTCACATCGTAGCTTCTGCAACCG	360
DB	4928	TATTCCTCTGCTGCGCAGACCTTTGGCAAGCTGCGGCATGTGCGAAGCTGTTGACGCGG	4987
QY	361	GGACTGATCTTTGCTGCTCCGATGCGACACCTTTCCAGCGGCAATTAGACCAATTTGCGCG	420
DB	4988	GGGCTGTGTTGTCACGCAACGCGACGCTTACGAGCGGCGCATCGACGCGGTGCTGCCG	5047
QY	421	GACGAGTCCCGCAATCTTCACTCGAGGCGAATTTGGCGGGCGGCGCAAGGTGATTTT	480
DB	5048	CCCGAGACCGCGGTATGACGCGTGGCGTGGCCAGGTGCCGGGGGCGCGCAAGGCACTTC	5107
QY	481	GACAGCTCTGTGAGACGCTGTGTGGATTTGAGGCGAGATTAATGCTTTGCGGCAATGCGC	540
DB	5108	GCGAGCTCTGTGCGCGAGCGCGGTGGCGCGAGGCGGAGTGGCGCTTTGCGGCAACGCGG	5167
QY	541	CCGATACGATTCGCAAGTCTTGTTCACCTTGAGCTTACCAATGCGCTTAAGCGGTG	600
DB	5168	CCGAGACGATGCGCAAGTTCCTTTCACCTTGAGGCTTCACCAAGTTCGCGCAAGGCGGTG	5227
QY	601	CCGACTACTCAGGAAATGCTTGGCGCAATACAGAGATCTTCTGCAACTTTCCGGTTT	660
DB	5228	ATCACCAACCGAGGATGCTTGGCGCAACAGCAATGCTGCTGAGACCTTTCCGGTGG	5287
QY	661	TTTGGTGAAGAGCGCGGTGCTGTGTGACTGCTGTTCCGTGGAAACCAACCTTCCGGCGC	720
DB	5288	TTTCGGTGAAGAGCGCGGTGCTGTGTGACTGCTGTTCCGTGGAAACCAACCTTCCGGCGC	5347
QY	721	AGCCACAACATGCGGATGCTGTTGTACAAGGCGGCGAGCATACCTTGAACGAGGTAATA	780
DB	5348	AGCCATTAAGTGGAGCATGCTGCTTACCAAGGCGGATCTTTTATCCTGGAACGAGGCAAG	5407

```

Oy 781 CCAACCGCCGAGGGTTGCGCGAGACGCTTCGCAATTGACGCAATCTCTCCACTGCG 840
Db 5408 CCAACCGCCGAGGGTTGCGCGAGACGCTTCGCAATTGACGCAATCTCTCCACTGCG 5467
Oy 841 TACCTCACTGTCGCGAAAGGCTGGAGAGAAATGATGGGTCCTTGAAGCAGACATACC 900
Db 5468 TACCTCACTGTCGCGAAAGGCTGGAGAGAAATGATGGGTCCTTGAAGCAGACATACC 5527
Oy 901 CTGCGGGAAGCTTTCTTCCGTCGATGACGCTTTCTTCCGCGGCGCTGGGTTCTCG 960
Db 5528 CTGCGGGAAGCTTTCTTCCGTCGATGACGCTTTCTTCCGCGGCGCTGGGTTCTCG 5587
Oy 961 CAAAGGATCTGGAGATCGTTTGACACCGGGTCTGTGAACAGACATGTTGAGCGCATTCG 1020
Db 5588 CAAAGGATCTGGAGATCGTTTGACACCGGGTCTGTGAACAGACATGTTGAGCGCATTCG 5647
Oy 1021 ATGATGCGGGGTCTGGGATGACGAGACATGCTCTTCTGCACTTTTACACCGACCG 1080
Db 5648 ATGATGCGGGGTCTGGGATGACGAGACATGCTCTTCTGCACTTTTACACCGACCG 5707
Oy 1081 CTGTCGATGCTGCTTTCATATTTGGGCTGCGACGCTGGCTGCGAGGTCGATCTTCG 1140
Db 5708 CTGTCGATGCTGCTTTCATATTTGGGCTGCGACGCTGGCTGCGAGGTCGATCTTCG 5767
Oy 1141 GTGATGAGGAAATTTGAAAGGCGCTTTCATGTCGCGACGTCATGAGCGGCTACTGCGT 1200
Db 5768 GTGATGAGGAAATTTGAAAGGCGCTTTCATGTCGCGACGTCATGAGCGGCTACTGCGT 5827
Oy 1201 GCTCTGAAACAAATGCGCCCAAGCGTTTGACAGAGAGAGCTATTAATGCTCTCGGTATGCC 1260
Db 5828 GCGGCGCAGACAGCGCGCGAGGTGTTGACGCCCAAGGCTTCTACTGCTCGGGGATGCG 5887
Oy 1261 ATGAAATTTGGCAATCTCTGCGCATCTTCAGAAAGTGTATGTTTGAACGCTCGAATTGCT 1320
Db 5888 CTCAACCTGCGCCAGCGCGCGCGAGACCGGCACTGGGCTGTATGTTGACGGGCGCATGCC 5947
Oy 1321 GAAGACTTCAAGCTGTCCTCAGAGGGTATTTGTGACGCTGAGGCGCATGCGCGAGCG 1380
Db 5948 GAAGACTTCAAGCTGTCCTCAGAGGGTATTTGTGACGCTGAGGCGCATGCGCGAGCG 6007
Oy 1381 GTTCTGAAAGCGCGCTTCTTACGTCCTGAGAGTGTGCTGCTGCTGTAATGCTGAATGC 1440
Db 6008 GTCTCGAAGCGCGCGCTTATGTCAGAGACCTGGTGTGCGCGCGCGAGATGTC 6067
Oy 1441 CTGGAATGCTGCTGCTTTCGCGCTCTTCTGACATGCGCTGCTGTGTGGGCTAGGAAA 1500
Db 6068 CTGGAATGCTGCTGCTTTCGCGCTCTTCTGACATGCGCTGCTGTGTGGGCTAGGAAA 1500
Oy 1501 GAGGCGTCGAGCGCGCGAGGTGCTTGCAGTGAAGCGGCTGCGGCTGCTGCTGCTGAATGC 1560
Db 6128 GAGGCGTCGAGCGCGCGAGGTGCTTGCAGTGAAGCGGCTGCGGCTGCTGCTGCTGAATGC 6187
Oy 1561 CTCAAAAGACTCAATCGAAGAGCACTGGCAATGCGAGTCATGTCGATGAGGCTC 1620
Db 6188 TTGCAAGTCTGATGCTGAGGCGCAAGCGGCAATGCGCGCTGAGTGTGATGCGCTG 6247
Oy 1621 CTGATAGCGCGCGCTGATGATTAAGGCGAGGCTCACTGACAAAGGCTGATTAACAG 1680
Db 6248 CAGGTGAGCGCGCGCTGATGATTAAGGCGAGGCTCACTGACAAAGGCTGATTAACAG 6307
Oy 1681 CCGGCTGTTTGGCAATGCGCGCTGCGGCGAAAGTTGATGCGCTGATGCGTGAAGATCAA 1740
Db 6308 CCGGCTGTTTGGCAATGCGCGCTGCGGCGAAAGTTGATGCGCTGATGCGTGAAGATCAA 6367
Oy 1741 TCCATGCTGCTGAGCAGAGCCACACTGTGA 1770
Db 6368 TCCATGCTGCTGAGCAGAGCCACACTGTGA 6397

```

RESULT 5
 AEO16866/c 310029 bp DNA linear BCT 09-OCT-2003
 LOCUS AEO16866
 DEFINITION *Pseudomonas syringae* pv. *tomato* str. DC3000 section 11 of 21 of the

```

ACCESSION complete genome:
VERSION AEO16866
KEYWORDS AEO16866.1 GI:28853240
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
PUBMED
REFERENCE
AUTHORS
FEATURES
source
1. 310029
/organism="Pseudomonas syringae pv. tomato str. DC3000"
/mol_type="genomic DNA"
/strain="DC3000"
/db_xref="taxon:223283"
/feature="pathovar: tomato"
204. 1664
/locus_tag="PSP02809"
/locus_tag="PSP02810"
204. 1664
/locus_tag="PSP02809"
/locus_tag="PSP02810"
/locus_tag="PSP02810"
1661. 2635
/locus_tag="PSP02810"
/locus_tag="PSP02810"
/codon_start=11
/transl_table=11
/product="peptide ABC transporter, permease protein"
/db_xref="GI:28853242"
/transl_table="MIGALRLRLOSLLVILMTVVVFGVIAIGNPMIDLVGEDLNO
ABRLQADLGLDQPLMOOYIFLKGAMGNIGOSVYVEDAMRLIOLPAPPELAF
SALFLAVVIGVPLGNFAGMTPDHPLSRMLMAHSIVSPSPAPVALMILNTEITLGH
LPSGREGTEWLVGVSWLITDGLQHLPLPALNLAFKISLVLRLTRAGVRLVDOE
FVFEARAKGISLPMRVCMCHVMRTMIIPVTVLAMELGISTAVAVTESIFAMPAGKLI
IDSINMLDRPVVAVIYLVVIFVILNLIIVDLVYLDPVRVIRASR"

```


gene 2632. .3552
/locus_tag="PSPRO2811"
CDS 2632. .3552
/locus_tag="PSPRO2811"
/codon_start=1
/transl_table=11
/product="peptide ABC transporter, permease protein"
/protein_id="AA056311.1"
/db_xref="GI:28853243"
/translation="MNVIKVAIPANTLLAQSPRRIVDFMSSSTAVVGLLLLVIT
LLAAPWIVQNPVDLMQNLNMDAMPKSGTVMYGTGDSGQSDLVASAIYQ
LRISLWVIGSALIAAVITGLGVSAYVAGWDLMLKLVLLSPYITLMAIYLA
WKGKGVNMLTLLLEMAVYARTAGQALTESRBYVDARQGVGPURIVYGHLP
NCLPLIVIGALQIARAITLEATLSFLGQVPTSPSLGLLIANGFQWMLSEYWISL
FGLALITIVIAINLIGRLRDVLPRLQR"
3549. .4562
/locus_tag="PSPRO2812"
CDS 3549. .4562
/locus_tag="PSPRO2812"
/codon_start=1
/transl_table=11
/product="peptide ABC transporter, ATP-binding protein"
/protein_id="AA056312.1"
/db_xref="GI:28853244"
/translation="MSLPVSRVTSSEPTLEVRDICTSFYTRAGVLPVARNVSLRQ
PRLIGLVESGSGKSVTPSILGLVDEGRISGGVLFKGRDLTKLSTGQREIOGN
RAAMIFQDPMATLPVLRVDTQMIEVRAHSTIGKREAREHARSRLTALNGISPEERL
RIYPHOLSGMRORVAIALAHAPDLITADEPTTALDVITQQLISEYOKLVEEGGT
SLIMTTHDLSVAGLADVAIVAGRIVEGSGVADVLDRPOHYTOGLIDSLPSRRKR
GRLRQIPGMARDLLSMRPGCAFAACAAAPACADPKPREILPGHTYRCYPPGAGA
DVQ"
4552. .5565
/locus_tag="PSPRO2813"
/note="This region contains an authentic point mutation,
causing a premature stop, and is not the result of a
sequencing artifact; peptide ABC transporter, ATP-binding
protein, authentic point mutation; similar to GP:15158466;
identified by sequence similarity; putative"
5618. .7183
/locus_tag="PSPRO2814"
CDS 5618. .7183
/locus_tag="PSPRO2814"
/codon_start=1
/transl_table=11
/note="similar to GP:15140494, and SP:P23847; identified
by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="peptide ABC transporter, periplasmic
peptide-binding protein"
/protein_id="AA056313.1"
/db_xref="GI:28853245"
/translation="MKPLVRSLLASALVLAAGSVSNQDRLIGVADPVSSLDLPQNLNNA
GGRSVLHAFESLVSRRDDKTLPGAKLSWKYTDIDTWFEALREDVYKMOGTPPLTADL
VSPFERARVPGSVASVAGAMRVSVKAKDHTLLIKRLPANLILPVDVSIYYSR
HGAQAASSADVNSGKALIGTGPYRPVSPICGORTIFARNDSWGAKPTMDKIDFRITA
NAAKRTALLAGDVVIDRVSPDVERLKRITSVNFAYQGLALTIQSPFRGPEP
IDNAGKPLAENPLDLDVRVKALSLIANPALDERIMQGVTEANMOMFANPENG
INIVPDVQKADLLAQAGFPBGFLTVVPGRYEQAEVWQAVQAFYTRIGVKKOL
EVLPMVAVGAKNKELAISVLAMNGTSEAAVLTNLTVDSSSGQASWAGHSN
PLVDKALADSTAEPEAKRKRLIEDSVGVSDVVGIIPLFHYQNIWDARKKLKVEPLV
SPRTAAITWTEQP"
7255. .7776
/locus_tag="PSPRO2815"
CDS 7255. .7776
/locus_tag="PSPRO2815"
/note="similar to GP:15162228; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA056314.1"
/db_xref="GI:28853246"
/translation="MTTERSNTPDULLHLGIFPGNPLTVYHARPKVASATQSGAL
FDPALADNLPISERLWVAIVATQUSAOPTLTATYVYIQQLREVGVADSVLADVANRID

gene 11EDPRLAAILFRTLLIESPVHGDSALQALQLOGSLTAETVTLAQLIAPLSYQVRL
AAGLTAIQSAGAA"
CDS 7776. .8366
/locus_tag="PSPRO2816"
7776. .8366
/locus_tag="PSPRO2816"
/note="similar to GP:15162229; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="alkylhydroperoxidase Ahpd domain protein"
/protein_id="AA056315.1"
/db_xref="GI:28853247"
/translation="MSELIEIGFTNQVLGWKAMLPYVDLNATAEQTALESHPQA
RTSDYVLTAAHHPDLIRQSQAFNAIMVAPGGISAEELASTVSRINRCYCASH
AQRPOLAKRNDVIAOVFADPATGTAEKIAVOFADLTLEPASINAAHIALREQ
GLDDAQVLDLJHAIAIFMANRLMLNIGSPVPEAE"
8429. .8614
/locus_tag="PSPRO2817"
CDS 8429. .8614
/locus_tag="PSPRO2817"
/note="similar to GP:15158870, and GP:15980012; identified
by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved domain protein"
/protein_id="AA056316.1"
/db_xref="GI:28853248"
/translation="MAECHPSLFMVCARVSATVTFEPGARTAMTHPLGQTLIVTAG
AQVSDAEVASFAPQP"
complement (8591. .9982)
/locus_tag="PSPRO2818"
CDS complement (8591. .9982)
/locus_tag="PSPRO2818"
/note="identified by match to PFAM protein family HMW
PF01569"
/codon_start=1
/transl_table=11
/product="fatty acid desaturase"
/protein_id="AA056317.1"
/db_xref="GI:28853249"
/translation="MYILVYGLLYPLVYHPVKSRLVQPSGVDAHVPLLPPTLPYL
STLVMPVLVYKRGSSWILPFTFAGALAAAGLCVLEFPMILIRPEIGAMMDLKY
QUDAPLAPSPSHVALPVAISVAMGLRQKMAVAVMSVIMLVMTVTGQHVFDWA
YGAAGVACGAATLIFKRCAYDLRTLSALLEMILCIVLRVALYLDWRFYLLAMLI
VBAROHALFVLXDAHYNLTRRSRNDPLINLAIGVPLVIEFYRPLADRHQHLG
TQDEBRRLYRQWRPQPLSAKLLRLQDLGDLVNTLRNIAVKAAGAPKPTR
PLISALVIMLVACLVQOCISREVGLMLMFIPLVYGTLLQKRSIAESGGGV
TGMQEWYAMRVGMLGRFTWPIYHINHLQHRTRASIFMHALPSAVGKDEKLMPERA
LPALMWGSLRRY"
10166. .10954
/locus_tag="PSPRO2819"
gene
Query Match 54.5%; Score 965.2; DB 1; Length 310029;
Best Local Similarity 71.6%; Pred. No. 6e-221;
Matches 1267; Conservative 0; Mismatches 503; Indels 0; Gaps 0;
QY 1 ATGCGTTCTCTGAGGCGCTTCCCTCCCGGGTGAATTCCTTGAGGCTCTCGAGCAT 60
DB 172246 ATGCCAGCCCTGGAACCGTTGGCGCCGTCGCCGATCGCTTCTGACACGGCTGGTGCAC 172187
QY 61 TGGGCTAAGACCCGTCGAGAACAAACCTGCTGTCGACGAGGCGGCAATGGGGAGATG 120
DB 172186 TGGGCGAGCGCTCGTGCAGAACAAACCTTATTTGCGCGCCGTGACAGCGGTGGTGGCTGG 172127
QY 121 CGTCGATAGCTACGCGGAATGTTCCCAACAAGTCCGCGGCACATGCGACAGAGCTTGCTT 180
DB 172126 CGCAAGTACGTTACCGCCGACATGCTCACCGAGTTCGGCCATTTGCCAGAGACCTGCTG 172067
QY 181 CCTTACGACATACGCGACAGCGTCCGCTGATTCGCTCTGGAATATGACCTGGAACAT 240
DB 172066 GCTTACGGGCTTTCCGACAGAGCGGCACCTGCGGCTGCTGTCCGGTAAACACATCGAGCAT 172007

QY 241 CTTGAGCTGGCATTGGGGGCTATGATCGGGGCAATTCCTATTTGGCCGGTGTCTCGT 300
 DB 172006 TTGCAATTCCTTTGGGCGCTATGTACCGCGATATCTTATTTGCCGGGTGTCAACGGCT 171947
 QY 301 TATTCACTGCTGTGCGAAGATTTGGCGAGCTGGCGACATCGTAGGCTTTCTGCAACCG 360
 DB 171946 TACTCGGTTCTGTGCGAGACTTCCGCAATTTGGGAGATGTCTGTAGAGTTTACAGCA 171887
 QY 361 GGAATGCTTTTGTGCTGCGATTCAGACACCTTTTCAGGCGCGCAATTTGACCATTTCTCGC 420
 DB 171886 GGCTCGTGTGTGTGTACAGCGATTCGCGCCGCTTCCAGGCGCGCATCGATGTGTATACCG 171827
 QY 421 GACGAGCTGCGCCGAATCTTCACTCGAGGGGAATTTGGCGGGGGCGGCAAGTGTGATTT 480
 DB 171826 GCTGACACACCGGTCATTACGATACGCGACAGTTGAATGGCGGCTCATTTGAGCTTT 171767
 QY 481 GACAGCTGCTGAGACAGCTGTGTGGATTGAGGCAATATATGCTTTGGCGCACTGGC 540
 DB 171766 GCCAGCTGCTTGAACAAACCGGGGGGCGAGGAGCGACCGCTGCTTCAATGGCCAGCGGG 171707
 QY 541 CCGGATACGATTTGGCAAGTTCTTGTCACTTCTGTGCTTACCAATGCTTGAAGCGGTG 600
 DB 171706 CCGGACAGCAATTCGCAAGTTCTTGTCACTTCTGTGCTTGAACCAATGCTTGAAGCAAGT 171647
 QY 601 CCGACTACTGAGGGAATGCTCTGCGCCATCAGGAGATGCTTGTGCAACTTTCGGGTT 660
 DB 171646 GTTACACTAGGAGCATGTGTGTGCGCAATACAGATGCTGCTGCAAACTTTCGGGTTG 171587
 QY 661 TTTGTGTAAGAGCGCGGCTGT 720
 DB 171586 TTGGGGGAACAGCGCGGCTGT 171527
 QY 721 AGCCACACATCGGACGCTGT 780
 DB 171526 AGTCATATACCTCGATATCGT 171467
 QY 781 CCAACGCGCCCAAGGCTTGT 840
 DB 171466 CCGACTGT 171407
 QY 841 TACCTCACTGT 900
 DB 171406 TACCTCACTGT 171347
 QY 901 CTGCGGGAAGCTTTCTGT 960
 DB 171346 TTGCGTGAGGAGATTTCTGT 171287
 QY 961 CAAGGAGCTGGGATGTTGT 1020
 DB 171286 CAAGGATGTGGGACCGCTGT 171227
 QY 1021 ATGATGCGGGT 1080
 DB 171226 ATGATGCGGGGCTGT 171167
 QY 1081 CTGTGATGCTGT 1140
 DB 171166 CTGTGATGCTGT 171107
 QY 1141 GTGTGATGGAATTTGAAGGGGCTTTCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1200
 DB 171106 GTGTGATGGAAGCTGTGAAGGGGCTTTCGCGGCGCGCATGTGTGTGTGTGTGTGTGTGT 171047
 QY 1201 GCTCTGGAACAAATGCGCAAGCTTTCAGAGGAGGATTTACTGCTCGGTGTGTGTGTGTGT 1260
 DB 171046 TCGCGGAGCAAAACCGCTGT 170987
 QY 1261 ATCAATTTGGCAAGATCTTCCGATCTTCAAGAAAGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1320
 DB 170986 ATCAAGCTGTGCAAGCTTCTGT 170927
 QY 1321 GAAGACTTCAAGCTGTCTCAAGGGGTATTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1380

DB 170926 GAGGATTTCAAACGT 170867
 QY 1381 GTTCTGGAAGGCGGCTTTAAGTCTGTGACGTAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1440
 DB 170866 GTGCTGGAAGGCTGTGCGGTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 170807
 QY 1441 CTTGATTTGCTGT 1500
 DB 170806 CTGGGGGCTGT 170747
 QY 1501 GAGGCTGTGAGCGCGGAGT 1560
 DB 170746 CAGCGAGCATGT 170687
 QY 1561 CTCAACGACTCAATCGAAGAACATGT 1620
 DB 170686 CTGGCACGTTTGAATGAAGGGGCAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 170627
 QY 1621 CTGATACGCGCGCTGT 1680
 DB 170626 CTGATACGCGCTGT 170567
 QY 1681 CCGGCTTTTGTCAATGT 1740
 DB 170566 CTGT 170507
 QY 1741 TCAATGCTGT 1770
 DB 170506 TCAAGGCTGT 170477

RESULT 6
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

AE016786/c
 301214 bp DNA linear BCT 12-DEC-2002
 Pseudomonas putida KT2440 section 13 of 21 of the complete genome.
 AE016786 AE015451
 AE016786.1 GI:26557030

Pseudomonas putida KT2440
 Pseudomonas putida KT2440
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 Pseudomonadaceae; Pseudomonas.
 1 (bases 1 to 301214)

REFERENCE
 AUTHORS
 Nelson,K., Paulsen,I., Weinel,C., Dodson,R., Hilbert,H., Fouts,D.,
 Gill,S., Pop,M., Martins Dos Santos,V., Holmes,M., Brinkac,L.,
 Beanan,M., DeBoy,R., Daugherty,S., Kolonay,J., Madupu,R.,
 Nelson,M., White,O., Peterson,J., Khouli,H., Hance,I., Lee,P.,
 Holtzapple,E., Scanlan,D., Tran,K., Moazzaz,A., Utlarback,T.,
 Rizzo,M., Lee,K., Kosack,D., Mosti,D., Medler,H., Lauber,J.,
 Honesel,J., Straetz,M., Helm,S., Kiewitz,C., Eisen,J., Timmis,K.,
 Duesterhoft,A., Tummeler,B. and Frazer,C.
 Complete genome sequence and comparative analysis of the
 metabolically versatile Pseudomonas putida KT2440
 Environ. Microbiol. 4 (12), 799-808 (2002)

TITLE
 JOURNAL
 REFERENCE
 AUTHORS

TITLE
 JOURNAL
 REFERENCE
 AUTHORS

TITLE
 JOURNAL
 REFERENCE
 AUTHORS

FEATURES
 source
 1. 301214
 /organism="Pseudomonas putida KT2440"
 /mol_type="genomic DNA"
 /strain="KT2440"
 /db_xref="taxon:160488"

```
gene
CDS
    complement(69..1007)
    /locus_tag="PP3196"
    /protein_id="AA068805.1"
    /locus_tag="PP3196"
    /note="similar to GB:X79440, SP:Q16798, and PID:535012;
    identified by sequence similarity; putative"
    /codon_start=1
    /transl_table=11
    /product="modulation protein d1"
    /protein_id="AA068804.1"
    /db_xref="GI:24984828"
    /translation="MRFHLDLNLVALDILIEQNITPAERLHMTGSGTSGVRL
    RFFEBELLVQGRKQPTPYALELAPREVALTIRASITAPVPDPASRKHFLV
    ASDYLISVFAQVTKIHOEAPQITPEMIGPGDSSAELMRGEVMMIVPERYLIEGH
    PAOLLEEDHVCVVMRDNPEVEGHLTLBOYMOGHSVSGRNRHLSDIMFNNOYGF
    NRRLVITNDPNTLPOLLIGTVRVATMRRLAELVAGYLPRLRPPVAKIPVWRERML
    MRRSMDGDPMHMRKRISDVIGCASTQPDALRATP"
    1284..2225
    /locus_tag="PP3197"
    1284..2225
    /locus_tag="PP3197"
    /note="similar to GB:X54232, SP:P35052, and PID:31847;
    identified by sequence similarity; putative"
    /codon_start=1
    /transl_table=11
    /product="glyoxalase family protein"
    /protein_id="AA068805.1"
    /db_xref="GI:24984829"
    /translation="MWIIGLDALVGVNDIQAACDLRDYGLDVALVDLDDGGRFALD
    GRSIIRRGDDDELPPALGPTTSRLRTYGVASSTDLAIADBLGDRERYRDSAGL
    HFFDLGFAIAFOVTCRPPFDAADLTNAQSAPOKINQLGIAPMPALPRLSHV
    YVPPDAKGEAIAFYRDLGFVTDSFIVGPFMTKATAMDDHCLFMIOTPPHOGCHF
    TFMGSGTEVLLAGTFFLOQGWTSFMGPRHLGSMVFWNSPLCGHIEYADMDKH
    DDAWAAARQAPMSADNSQLFLPASREKAPSGPAKQG"
    2238..2576
    /locus_tag="PP3198"
    2238..2576
    /locus_tag="PP3198"
    /note="identified by match to Pfam protein family HMM
    PF00355"
    /codon_start=1
    /transl_table=11
    /product="ferredoxin, 2Fe-2S"
    /protein_id="AA068806.1"
    /db_xref="GI:24984830"
    /translation="MELCRDLDELSEGOARGDPQGRGADSLFALRYLGQVRYRNLCP
    HLMVPLEYRKRLSADGQWICYAHGARFTPEDDICVHGPKRGALDLDHKEAGWM
    LILPLEQDR"
    2716..3840
    /locus_tag="PP3199"
    2716..3840
    /locus_tag="PP3199"
    /note="similar to SP:P23262; identified by sequence
    similarity; putative"
    /codon_start=1
    /transl_table=11
    /product="monooxygenase, putative"
    /protein_id="AA068807.1"
    /db_xref="GI:24984831"
    /translation="MNTVNVNLVGGGIGSLCAALARRKGTAVDLVELKTEWTVYG
    GIIQGSNVREMRGLLEAYLDAAAFEDVALYTBEGQOLARI PQGRLAGPQYPANV
    GISRALHKLSETAIELGTVALGTSVESFEDDSDADVLPFGDSQGRGLMVGADG
    LRSKRVSLFGDRYCPRTGQSVRYNFPRAAGIDHLANQSGSGNAGVPLADLDLY
    LPTTSHPCNPMPPEPALADGRLORLAGFGLIGELREOIVDNAGVYKLEAVFVDE
    PMYCGVILLIGDAHAHTPHLIGGAGMALIEDAVILAEELTAGGSLDAQLORFNGRFE
    RKFTISEKSLAGVKETIQRPSTDRIGLVKRMLETTAQPI"
    3885..5597
    /locus_tag="PP3200"
    3885..5597
    /locus_tag="PP3200"
    /note="similar to GP:3184126, and GP:3184126; identified
    by sequence similarity; putative"
    /codon_start=1
```

```
gene
CDS
    /transl_table=11
    /product="conserved hypothetical protein"
    /protein_id="AA068808.1"
    /db_xref="GI:24984832"
    /translation="MPTPHNRKKEETPMWLRKGAAPRLPALKPLALVALAASQA
    QAEPDGTNPDLRVRMDNTLTKTSAMRITDASSKLTBQGTALNLDGDRNFDKGLISN
    RLDLSELDITVYNNVGARVSAAWDDVYNGSTDNNDPGRANVSVDYAFDDTRKL
    HGKRGHLDALFPGKTEIGDMPLSGRLGVAQMGESLFFGNNGIAGMAPYDVVVGGL
    SVYNPQFKELIRPVQOISGLOLTPDVSVAQYQFEMEANRLPAAGSYSTDDFFGDLG
    AARMFVGAPVPPSGGOLIAFHGNDKOAKDSGGGQMVRRTERVTDGLAIRPHDQSP
    QIANRPNRANLIDPDGSKAGEFYWYVDEGIEARGASPTTVGAPNLAGELSTRNQDIA
    STNQHALAVGEALIDNDNPVYATGRTLHANVMSLSENPFAAESSEFGEIANNRVN
    SVSKADADVPADDDATSLRVPFPMRQIIPGLDLSVPIGASVTNGASMAAGTGF
    TTHGGILNLTGLNGVNLMTSLDITVTHYVGPENTFLDADNHYTFEQLKDRDIAFTV
    STRP"
    7034..8095
    /locus_tag="PP3201"
    7034..8095
    /locus_tag="PP3201"
    /note="identified by match to Pfam protein family HMM
    PF02012"
    /codon_start=1
    /transl_table=11
    /product="BNR domain protein"
    /protein_id="AA068809.1"
    /db_xref="GI:24984833"
    /translation="NMKMLYGMWLLAACGLPPAFAGDSYDVLAQPALHGPOLRAVLQ
    DVARADARLVAWGERGVLLSDNGATWMAQFAVPGVLTAVQFIDARNGNAVGHAG
    VTLHSDQDGEHNSVOLDDGRKALALVOAAEAAADRRRLAAARLLADGADKRLALNF
    TTRAGNLVYGAIGVAMRTDGGSRNQSEIGRLPNRGHLHYALARGNVIYAGEGL
    LTRSDNGEYHALGEPDGSYFAATVLPDGLLJGLGSGSFPASDDGSDFOALANF
    TPASINGLRVAGGOLLLANQAOMLRSLGSLGAAPAVAVSDGLPLTAATTAADGSIVA
    VMAARARLRLEPSTTTD"
    8108..10588
    /locus_tag="PP3202"
    8108..10588
    /locus_tag="PP3202"
    /note="similar to GP:9949187; identified by sequence
    similarity; putative"
    /codon_start=1
    /transl_table=11
    /product="transporter, putative"
    /protein_id="AA068810.1"
    /db_xref="GI:24984834"
    /translation="MNSAPSTGPIDDPDSAGSLERALFNHRVLLVLLCLATLL
    GNOARLTLNASEFEKMIIPDHPIFNHYLEHROBELGNAVRIAVANPBGSIYDKOYL
    HSLQOLNDVNLTPVDPRAMSKIMTPSPSRMTGPEEGEPVDPVDGAAASIEA
    LKRNVERSEITQOLVAFDGRSSTIVYVPLLEKTPDQALDLYTPAHLETLRERFOAG
    VEIHITGFKAVVDLMDGLRQILFFPAALAIITAVLVITWTRCVSRITALLVCSLAV
    IWQGLPLGLYALDVPYSVLPVLFVAIQMSGAGQKMGIMODIGMRHLVAARFTF
    RRLFLAGLTALLCDADVFAVLMIOIOVODLAVIASLGVAVLIFNLILPLVLSYV
    GVSARBARSLRAEAEASGAKHVMRPLDFTERRRMAACIAVALMAAGSYAVSL
    HLKVDLDLAGABELRADSRYNMDNAFVTRHVGASSDVPAVMRTAPGCGSAGTTLKH
    DLDLDQRLGQVSDTNSLALNLRVYVGLSGSPKRYDLYVNNQATLNMVTAQAPRGL
    YNDGSLTLTVAYLTDHKADTLARVVDSDAQAANDSDASFAEASAGABATNQ
    VVKQNRDMLMNVGAVIVLCVTRSMKAVLCVLPVLSIILCEALMVALGIGIKVQ
    ATLPIVATGCVGIDVATVMSIVLAQOGASLSQAYVRALLFTGKVMLEGTITLAI
    GVGWTIPSPKIFQADMVGLIAFMFVNMWAGLILPALAYFLPRLKACTEQLALPAH
    EHTRRERSGMEQASFRPLDEVBGR"
    10578..11768
    /locus_tag="PP3203"
    10578..11768
    /locus_tag="PP3203"
    /note="identified by match to Pfam protein family HMM
    PF04171"
    /codon_start=1
    /transl_table=11
    /product="major facilitator family transporter"
    /protein_id="AA068811.1"
    /db_xref="GI:24984835"
    /translation="WVAERTAGSAQALILPSCCLPYGAVLIAPVLPBMHVAFAETP
    GVAVLPVALTIPALVITALLAPLAGTADRVRRALLIASMLIYSCGLPLPMDSLV
    LTVASRAGIGLAEAGIMTCCTTLMGDFDGGRRARIPALOMVVTSLAALPFGVGGAL
```

```

      GESDRTPTLVAVGALCLPLMLLLMEPRACOAASBASVDSRFPMAALPLVLTVL
      AGVSLFIVPAQVAVIOLHIDAPROVGLAMGANOIGVLAGALAPRLALPAGRLIA
      MGRATRGGLGGLMALSHAPVPLVAVLINLGLVLLPITITVMOQVGDGRATG
      GFTSALFAGEFVSPLVLALTAGVATPLPALILVALIGULLAPVCLSLMRHRAVTV
      AGAR"
      gene
      11926..12513
      /locus_tag="PP3204"
      /locus_tag="12513
      11926..12513
      /locus_tag="PP3204"
      /note="Identified by Glimmer2; putative"
      /codon_start=1
      /transl_table=11
      /product="hypothetical protein"
      /protein_id="AA068812.1"
      /db_xref="GI:24984837"
      /db_xref="GI:24984837"

Query Match      50.7%; Score 898.2; DB 1; Length 301214;
Beet Local Similarity 69.4%; Pred. No. 7.6e-205;
Matches 1221; Conservative 0; Mismatches 538; Indels 0; Gaps 0;

QY      1 ATGCGTTCTCTGAGGCGCTTCTTCCCTCCCGGTTGAAATTCTGAGCGTCTGAGCAT 60
DB      170844 ATGCAACCTGTCTGAGCCTGCGCGCTGCGCGCGCTGCTGCGCATCTGTGTCAT 170785
QY      61 TGGGCTAAGACCGCTCAGAACAACTGCGTTGCTGCGAGGCGCGCAATGGGGAATGG 120
DB      170784 TGGGCCCAAGTGGCGCGCGACCACTTTATGCGGACCGACGCAAGCGTGTGG 170725
QY      121 CGTCGTATCAGCTACGCGGAATGTTCCACAAGTCCGCGCATTCGACAGAGCTTGCCT 180
DB      170724 CGTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 170665
QY      181 CCTTACGAGATATCGGACAGCGCTCCGCTCTTATCGTCTGTGAAATGACCTGGAACAT 240
DB      170664 GGACTGGGCTTCACTGCTGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCT 170605
QY      241 CTTCAGCTGAGCATTTTGGGCTATGATGATGATGATGATGATGATGATGATGATGATGAT 300
DB      170604 CTGCAATATGCGCTCGCGCGCATGATGATGATGATGATGATGATGATGATGATGAT 170545
QY      301 TATTCATGCTGTGCGAAGATTTGGCGAAGCTGCGTCAATCGTAGGTCTTTCGCAACG 360
DB      170544 TACGCGCTGTGTCGCAAGACTTCGCAAGTGGCGCATGTCGCGAGTGTCTACACCC 170485
QY      361 GGAATGCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
DB      170484 GGAATGCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 170425
QY      421 GACGAGTGGCCGCAATCTTCACTCGAGCGAATTTGGCGGAGGCGCGCGAGTGTGTTT 480
DB      170424 GATTCGCTGCGCGTGAATCAAGCTGCGCGTGGCGAGTGGCGAGTGGCGAGTGGCGAGT 170365
QY      481 GACGAGCTGTGAGGAGCTGTGAGGATTTGAGGCAAGATATGCTTTTGGCGCAACTGGC 540
DB      170364 GACGAGCTGTGAGGAGCTGTGAGGATTTGAGGCAAGATATGCTTTTGGCGCAACTGGC 170305
QY      541 CCGGATACGATTTGCGCAAGTCTTGTTCATCTTGTGCTTTCACCAATGCTTAAAGCGGTG 600
DB      170304 CCGGATACGATTTGCGCAAGTCTTGTTCATCTTGTGCTTTCACCAATGCTTAAAGCGGTG 170245
QY      601 CCGACTACTGAGGAATGCTGTGCGCAATCAGAGATGCTTTCGCAAACTTTCCCGGTT 660
DB      170244 ATACACACCGAGGAGCTGTGCGCAATCAGAGATGCTTTCGCAAACTTTCCCGGTT 170185
QY      661 TTTGTGAGAGCGCGCGGTGCTGTGAGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGT 720
DB      170184 TTTGTGAGAGCGCGCGGTGCTGTGAGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGT 170125
QY      721 AGCCACAACATCGGATCGTGTGTGACAAAGCGGCGACGATCTTACGACGAGTAA 780
DB      170124 AGCCACAACATCGGATCGTGTGTGACAAAGCGGCGAGTGTTCACGACGCGCGGACG 170065
QY      781 CCAACGCCCAAGGTTGCGCGAGAGCTTTCGCAATTTGAGGCAATCTCTCCCATGCG 840

```

```

DB      170064 CCGACCCCGCAAGGCTTCCGCGAGACTTCCGCAATGTGCGCGAGATTTCCCGACGGCC 170005
QY      841 TACCTCACTGTGCCGAAAGCGCTGAGAGAAATTAGTGGTGCCTTTCAGCGAGACATACC 900
DB      170004 TACCTCACTGTGCCGAAAGCGCTGAGAGAACTGTTCAGCGAGACATACC 169945
QY      901 CTGCGCGAAGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
DB      169944 CTAGCGAGGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 169885
QY      961 CAAGGATCTGGATCGTTTGAACCGGCTGCTGCAAGCAGCATGCTGTGAGCGCATTCGC 1020
DB      169884 CAAGGATCTGGATCGTTTGAACCGGCTGCTGCAAGCAGCATGCTGTGAGCGCATTCGC 169825
QY      1021 ATGATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
DB      169824 ATGATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 169765
QY      1081 CTGTCATGCTGCTTACATTTGGGCTGCGACCGCTGCGAGGCTGCGAGCTGCTTCCG 1140
DB      169764 TTTGTCATGCTGCTTACATTTGGGCTGCGACCGCTGCGAGGCTGCGAGCTGCTTCCG 169705
QY      1141 GTCGATGGAATTTGAAAGGCGCTTTCATGCTGCGACCTGCAATGACGCGCTACTGCGCT 1200
DB      169704 GTGGGCGACAGCTGCAAGGCGCTTCCGTCGCGCGATTCATGCGCGGCTACTGCGCG 169645
QY      1201 GCTTCGTAACAATGCGCAAGCTGCTGCAAGGCTGCTGCAAGGCTGCTGCAAGGCTGCT 1260
DB      169644 TCCCGCGACAGACCGCGCGAGGCTTTCAGAGAGAGGCTTTCATGCTGCGCGAGCGCG 169585
QY      1261 ATCAAAATGGCAATCTGCGCATCTCCGATCTCCAGAAAGTCTGATGCTTGAATGCT 1320
DB      169584 TTAAGCTGCGCATGCGAGGAGCGCGAGGCTGCTGATGCTGATGCTGATGCTGATGCT 169525
QY      1321 GAAGACTTCAAGCTGCTGCGAGGCTTTCATGCTGCGAGGCTGCTGCGAGGCTGCTGCG 1380
DB      169524 GAGGACTTCAAGCTGCTGCGAGGCTTTCATGCTGCGAGGCTGCTGCGAGGCTGCTGCG 169465
QY      1381 GTTCTGGAAGCGGCTTTCATGCTGCGAGGCTTTCATGCTGCGAGGCTTTCATGCTGCG 1440
DB      169464 GTTCTGGAAGCGGCTTTCATGCTGCGAGGCTTTCATGCTGCGAGGCTTTCATGCTGCG 169405
QY      1441 CTGGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
DB      169404 CTGGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 169345
QY      1501 GAGGCGTGGAGCGCGCGAGGCTTTCATGCTGCGAGGCTTTCATGCTGCGAGGCTTTCATGCT 1560
DB      169344 GATGCGAGGAGTCCGCGGCTGCTGCGAGGCTTTCATGCTGCGAGGCTTTCATGCTGCG 169285
QY      1561 CTCAAAAGCATCAATGCAAGGAGCACTGCGATGCGATGCGATGCGATGCGATGCGATGCG 1620
DB      169284 CTGAGGCGTGTGAACCGCGATGCGCAAGGAGCAAGGCGGCTGCTGCTGCTGCTGCT 169225
QY      1621 CTGATACGCGCGCGCTGATGATTAAGGCGAGGCTGCTGCAAGGCTGCTGATCAACGAG 1680
DB      169224 CTGAGGCGTGTGAACCGCGATGCGCAAGGAGCAAGGCGGCTGCTGCTGCTGCTGCTGCT 169165
QY      1681 CCGGCTGTTTGGCAATGCGGCTGCGGCAAGTGTGATGCGCTGTGATGCTGTGATGATCAA 1740
DB      169164 CCGGCTGTTTGGCAATGCGGCTGCGGCAAGTGTGATGCGCTGTGATGCTGTGATGATCAA 169105
QY      1741 TCCATGCTGCGTGAAGAG 1759
DB      169104 GACGATTCGACCGCAAGG 169086

RESULT 7
PF0536324 4407 bp DNA linear BCT 09-JAN-2004
LOCUS PF0536324
DEFINITION Pseudomonas putida ech gene, vdh gene, fce gene and at gene
(partial).

```

ACCESSION AJ536324.1 GI:40788111
VERSION at gene; beta-ketoliolase; ech gene; fcs gene; feruloyl-CoA
KEYWORDS hydratase; feruloyl-CoA synthetase; vanillin dehydrogenase; vdh
gene.
SOURCE Pseudomonas putida
ORGANISM Pseudomonas putida
Pseudomonas putida
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1
AUTHORS Ficca A.G., Di Gioia D., Barghini P., Fava F. and Ruzzi M.
TITLE Identification of Pseudomonas putida strain T2 genes involved in
ferulate catabolism
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 4407)
AUTHORS Ruzzi M.
TITLE Direct Submission
JOURNAL Submitted (09-JAN-2003) Ruzzi M., Agrobiology and Agrochemistry,
DABAC - University of Tuscia, via C. de Lellis, snc - Viterbo,
01100, ITALY
FEATURES
source 1..4407
Location/Qualifiers
/organism="Pseudomonas putida"
/mol_type="genomic DNA"
/strain="T2"
/isolation_source="soil"
/db_xref="taxon:303"
/country="Italy:Lazio"
10..741
/gene="ech"
10..741
/gene="ech"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="feruloyl-CoA hydratase"
/protein_id="CAD60261.1"
/db_xref="GI:40788112"
/translation="MSPTLNREMRDVLETVEODADARVLVLTEGSGSAMTAGMDLKEF
REVDAPEILORIRRDASQWKLRFYSKPTIAVNGMCFEGSGPLVACDLAICA
DEATRELSEINNGIIPGNLVSXAMADTVGHRESLYIMGKTPDGOKAOMGLVNSVS
PLAOLRDEVVLVAQDLIDKNPVLPAKNGFKRCRLTWEQNEHYIYALDQAEIDPE
HRAQGLKQPLDCKSIKPELQAYKR"
782..2239
/gene="vdh"
782..2239
/gene="vdh"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="vanillin dehydrogenase"
/protein_id="CAD60262.1"
/db_xref="GI:40788113"
/translation="MPDVRGALLIGKKAASNGTFFRRBHPUSGEVSVVAAATLAD
ADAAVBAQAAPFAPMAALPGERRARILKAADALDRRDEFLMASETKAKAMTFN
VMAAVILREASMTTQIQSVYPSDVPSFAMALRQCGVVLGA PMNAPVILATRA
IAMPACGNTVVLKASBPVAVHGLIGVYLAAGLDGVAVVLCNAPHPAIVERLI
AMPAYRNVFTGSTHVGRIVGOLAARLKPAYLELGGKAPVLADADLEAAVADAA
FGAYVNOGQICSTERLIVDOAVADEFAVKIGEKVRLAAMLSOPDAPLGSILDGA
GERIALYNDAYGAKATVAGGERGTSVWQPVLDHYNAGMLYREESGPAYVYRA
SGDEBALALANDSERGLSAITFSRDTORLALAAQRLESGICHINGFTVDEADMPGG
VVASGSGFSGTAIDAFQTLRWVTVQHARHPYI"
2319..4175
/gene="fcs"
2319..4175
/gene="fcs"
/gene="fcs"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="feruloyl-CoA synthetase"
/protein_id="CAD60263.1"
/db_xref="GI:40788114"
/translation="MDCEFLSVAGRAPLRAVKIQGPQVQVQGGVLYLDPEVETLDAY

PARLMEILLHMAQQRSEQTFVARBAADQWQHTTYAAMIQVRRIATWLLQGLSAR
PLIILISNDLEHVOLATALVAGIPIYCPSPSPYSLISQDFGKLRIHVLYQLQELVAA
DAAYGAILDAVEFSPSINTLVIGGELGAPMSTPFAEVLTTVAQPCADQAPLADBLA
KELFPGSGTKLRAVTTQGMLCANQOMLOJEPBEAEVPLVDMLPNNHTGGSHN
VGLIVYNGSTLYIDQKPTPOGPAOTIHLKOISIPRAYLTPVGMELVKALREODAL
CRRFTFROMLFFFPAGAGLSQATMDRLDAVAVAHCGERIRIMAGLCTERTSPSTPTTG
PLSVAGYIGLPAPEGCEVKLVPAVDGKLEAFRGHVMVGRWSQOTQEAHDEGFCS
GDVAVRLCDPGNPOLGLMPDGRJAEDFKLSGVFVSVGPRQTVSSGSPYLDVVVGP
DECEGLIIFPRLPECRQLSGAQOSTPKVLSFVAPRCLASILTALNRGATGNASY
IRFAGLAAPAPRIDRGEITDKGSLNQRVVLQJLRHGLIDSLVAGPRLVADAP"
4172..4407
/gene="aat"
4172..4407
/gene="aat"
/codon_start=1
/transl_table=11
/product="putative beta-ketoliolase"
/protein_id="CAD60264.1"
/db_xref="GI:40788115"
/translation="MTARGQFRTFATALLLEGIRTPWVDINGAFGLVSPDILGKAGR
AVLQRRARLDAGBVDVSLAGSMAQADFADALLBRH"
ORIGIN
Query Match 43.1%; Score 763.2; DB 1; Length 4407;
Best Local Similarity 66.0%; Pred. No. 2e-172;
Matches 1136; Conservative 0; Mismatches 578; Indels 6; Gaps 2;
QY 11 TCGAGGCGCTTCTCCCTTCCGGGTCGAATTTAGGCGTCTCGACATTTGGGCTAAGA 70
DB 2431 TGGAAACGCTCGATGCTTACCCGACCGGCTATGAGCGCTTGCTGCATTTGGGCCAGC 2490
QY 71 CCGGTCAGAAACAACTGCTGTTCTGCCAGGCGGCAATGGGAATGGCGTGTATCA 130
DB 2491 AGGCGAGCAGACAGAGTGTGTGCGCGCGCTGCCAGCGGCGATGGCAGACATCA 2550
QY 131 GCACCGGGAATGTTCCACACGTCCGGCCATGGCAGAGCTTGCTTCTTAAGGAC 190
DB 2551 CTYACCGCCGCAATGTTCAGCGGTTCAGAGATCCACAGTGGCTTGGGCCAAGGCC 2610
QY 191 TATCGGCAAGAGCGTCCGCTTATGCTCTGGAATGACCTGGAACATCTTCACTGG 250
DB 2611 TGAAGCCGGAACCGCCGCTACTGATCTCTCGGCGACACACCTTGAGCACTGCACTGG 2670
QY 251 CATTTGGGCGTATGATGCGGCGATTCCTATTTGCCGCTGTCTCTTATTACTGTC 310
DB 2671 CCACGCGCGCGCTGATGCGGATTCCTATTTGCCGCTGTCTCTTATTACTGTC 2730
QY 311 TGTCCGAATTTGGGGAAGCTGCGTCATCATGTAGTCTTTCGCAACCGGACGTGCT 370
DB 2731 TGTCCGAATTTGGGGAAGCTGCGTCATCATGTAGTCTTTCGCAACCGGACGTGCT 2790
QY 371 TTGCTCCGATGCGACCTTTTCAGCGGCAATTGAGACATTTGCGGAGCAAGCTGC 430
DB 2791 TCGCGGCTGATGCGG---CCTAGCGAGGCGCATGATGCGGCTTTCGAACTCGATTA 2847
QY 431 CCGCAATCTTCACTGAGGCGGAATTTGGCGGCGGCGCACGCTGAGTTTGACAGCTGC 490
DB 2848 ACTGTGTGTGGGCGCAGGCGAGCTGCGGCGTGTGGCGAGACGCGCATTCGCGGAGTAC 2907
QY 491 TGGAGAGCGCTGTGGGATTTGAGCGAGATATGCTTTCGGGCAATGCGCCCGGATACGA 550
DB 2908 TGAACAACGTCGCCAACCTGTGCGGACAGAGCTTTGCGCGCTTGGCGCGGACAGCC 2967
QY 551 TTGCAAGTTCTTGTTCATCTTGTGCTTACCAAACTGCTAAGCGGTGCGGACTACTC 610
DB 2968 TGGCAAGTTCTTGTTCATCTTGTGCTTACCAAACTGCTAAGCGGTGCGGACTACTC 3027
QY 611 AGGGAATGCTTGGCGCAATGAGCAGATCTTTCGAACTTTCCGCTTTTGGTGAAG 670
DB 3028 AGGGAATGCTTGGCGCAATGAGCAGATCTTTCGAACTTTCCGCTTTTGGTGAAG 3087
QY 671 AGCGCGGCTGCTGTGAGCTGTTTGGCGTGAACCAACCTTGGCGGCGGAGCAACA 730

Db	3088	TCGCCGCGGTCTGGTGCAC	TGCGGTGCGGTGGAAAC	CACACGTTTGGCGGACGCCAACAG	3147
Qy	731	TCGGCATCTGTATTGTAC	CACGCGCGGCACAGTAC	TCTTGAAGACGCGTAAACCAACCGCC	790
Db	3148	TCGGATCTGCTGTAC	CAACGCGCGGCACGCTGTAC	TGATGAGATGGCAAGCGCACGCGCG	3207
Qy	791	AAGGATTCGCGGAGC	CGCTTGGCACTTGAGCGAA	ATCTCTCCACATGCGTACCTCAC	850
Db	3208	AGGCGTTTGGCCAGAC	CGCTTGCAACACTTCAGAG	ACATTTTCCCCACGCGGTACCTGACCG	3267
Qy	851	TGCCGAAAGGCTGGAG	AGAAATTAGTGGGTGCGCTTG	ACGAGACGATACCTTCGCGCAAC	910
Db	3268	TGCCGAAAGGCTGGAG	AGAACTGTGCAACACTTC	AGAGACATTTTCCCCACGCGGTACCTGACCG	3327
Qy	911	GCTTCTTCGCTCGC	ATGAACTGTTCTTTC	TGCGCGCGGCTGGTGTTCGCAAGGATCT	970
Db	3328	GGTCTTCTCAACG	GCATGAAGCTGTCTT	TTTTCGCGCGGTGCGGGTGTTC	CCAGGCGCATCT 3387
Qy	971	GGGATTCGTTTGGAC	CCGGGTCCGTAAACAG	CACTGTGTGAGCGCAATGGCATGATGAGCGG	1038
Db	3388	GGGACCGGCTTGA	CGCCGTGGCCGTGGCCCA	CTGGGCGGAAACGATTCGCGCATGATGAGCGG	3447
Qy	1031	GTCGAGGATGACG	GAGACTGCTCTTC	CTGCACTTTTACGACGCGACTGTGATGG	1097
Db	3448	GGCTGGGATATAC	CGAGACTTTCGCTGTG	CACTTTACACACGCGGCGCTGTGTGGTG	3507
Qy	1091	CTGGTTACATTGG	GCTGGCCAGCCCTGG	CTGGAAGCTCAAGCTCGTTCGCTGGTGAAGGA	1156
Db	3508	CCGGCTACATCG	CGCTGCGCGGCCCGGGGGTGG	AGATCAAGCTGGTGTCCGGTGAGCGGCA	3567
Qy	1151	AATTGGAAAGGCG	TTTCCATAGTCCG	CACATCAGACGCGGCTACTGGGCTGCTCTGAAC	1210
Db	3568	AGCTTCAGGCG	AGGTTTTCGCGGCGCCCA	CGTGAATCCGCGGATCTGCGCTCATCTGCAGC	3627
Qy	1211	AAAAATGCCAAG	CGTTTCGACGAGGAGGCT	TTTACTGCTCCGCGGTATGCCATCAATCAATGG	1270
Db	3628	AGACCCAGAGAG	CGCTTTCACACGAGAGGGG	TTTACTGCAAGGATATCGGTGCGCTGT	3687
Qy	1271	CAGATCTCTG	CCGATCTCCAGAAAG	GTCTGATGTGTGAACGATCGAATTTCTGAGAC	1330
Db	3688	GCGACCCCGGCA	CCCGCAGCTGGGTTT	GATGTGACGCGGCGCATTCGCGAGACTTCA	3747
Qy	1331	AGCTGTCTCA	AGGGGTATTGTCA	CGCTTGGGCCATTGGCGCACGCGGCGGTTCTGGAAG	1397
Db	3748	AGTTGTCTCT	CCGGGGTGTTCGTCA	CGCTTGGGCCCGATCGCACCC--AGGTATCTTCAG	3807
Qy	1391	GCGGCTCTTA	AGTCTCTGGAACG	TAGGTGTGCTGCTCCGATCTGTAATAGCTTGATATGC	1456
Db	3805	GGTGCGCTTA	TGTGTGTGATGTGT	GTGGGTGGGCGCGAGCCGAGCAATAGCTTGAGGTTTGC	3867
Qy	1451	TCGTGTTC	CCGCGCTTCTGCA	CTGCGCTTGTCCGGGCTAGAGAAAGAGCGTCCG	1510
Db	3865	TTATCTTCC	CGCGCTTGTGCGG	AAATCCGCAATGTCTCCGGGCTCCAGGCGCACTA	3927
Qy	1511	ACGCGAGGT	CTTTCACAGTAA	AGCCGTTTCGGGCTTGTGACTGGCTCAAAAGAC	1570
Db	3925	CGCCCAAG	GTCTGGCATCTG	CCCGCGTGGCGTGAATGCTGGCCAGCTGCGCACGCG	3987
Qy	1571	TCATTCGAA	AGCACTGGCAAT	TGCAATCGCATATGTGGGTAGGGCTCTTCATATAGC	1637
Db	3985	TCATC	ACCGGGGCGCGAC	GAGCAACGCACTATCCGCTTCGCCCGGCTTGGCCGCGAG	4047
Qy	1631	CGCGCTGAT	TGATTAAGGCG	AGGTCACTGACAGAGGCTGCATCAACAGCGCGCTGTTT	1697
Db	4045	CACGCGCAT	TGACCGTGTGTG	ATCATCCGACAGAGGCTCGCTCAACAGCCGCTGTGC	4107
Qy	1691	TGCAATGG	CGGTCCGCGAA	AGTTGATGGCGCTGTATCGTGG	1730
Db	4105	TGCAATGG	CGGCGCGCTGAT	TCGACAGCGCTGTATGACAG	4147

LOCUS	189050 bp	DNA	linear	BCT 11-APR-2003
DEFINITION	AL646077 Ralstonia solanacearum GM1000 megaplasmid, complete sequence;			
ACCESSION	AL646077			
VERSION	AL646077.1			
KEYWORDS	GI:17430642			
SOURCE	Ralstonia solanacearum			
ORGANISM	Ralstonia solanacearum			
REFERENCE	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Burkholderiaceae; Ralstonia.			
AUTHORS	1			
TITLE	Salanoubat, M., Genin, S., Artiguenave, F., Gouzy, J., Mangenot, S., Ariat, M., Billault, A., Brotier, P., Camus, J.C., Catolico, L., Chandler, M., Chisme, N., Claudel-Renard, C., Cunnc, S., Demange, N., Gaepin, C., Lavie, M., Moisan, A., Robert, C., Sarria, W., Schiex, T., Signier, P., Thebaud, P., Whalen, M., Wincker, P., Levy, M., Weissbach, J. and Boucher, C.A.			
JOURNAL	Nature sequence of the plant pathogen Ralstonia solanacearum			
MEDLINE	Nature 415 (6871), 497-502 (2002)			
PUBMED	21681879			
REFERENCE	2 (bases 1 to 189050)			
AUTHORS	Boucher, C.A.			
TITLE	Direct Submission			
JOURNAL	Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie Moleculaire des Interactions Plantes-Microorganismes INRA-CNRS, BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, LMGM CNRS 118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA URV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27, F31326 Castanet-Tolosan Cedex, Laboratoire de Genetique Cellulaire INRA, BP27, F31326 Castanet-Tolosan Cedex			
COMMENT	Christian.Boucher@toulouse.inra.fr http://sequence.toulouse.inra.fr/R.solanacearum.html.			
FEATURES	Location/Qualifiers			
SOURCE	1..189050			
	/organism="Ralstonia solanacearum"			
	/mol_type="genomic DNA"			
	/strain="GM1000"			
	/db_xref="taxon:305"			
	/plasmid="megaplasmid"			
	79..1500			
	/note="Rs49+ or RS06146			
	LONG DIRECT REPEAT 49 +			
	predicted by Homology"			
	/evidence=not_experimental			
	83..2971			
	/gene="RSP0175"			
	/note="synonym: RS04687"			
	83..2971			
	/gene="RSP0175"			
	/function="miscellaneous; hypothetical/global homology"			
	/note="Product confidence : putative			
	Gene name confidence : hypothetical			
	predicted by Codon usage			
	predicted by Homology			
	predicted by Framed"			
	/codon_start=1			
	/evidence=not_experimental			
	/transl_table=1			
	/product="PUTATIVE VGR-RELATED PROTEIN"			
	/protein_id="CAD17326.1"			
	/db_xref="GI:17430643"			
	/db_xref="GOA:Q8XTD8"			
	/db_xref="TrEMBL:Q8XTD8"			
	/translation="IMDKTLLIHLFAPAHRLVLAEGSGPIAEALAEVAMLGREALSELSELF			
	EMRWVLSANARFALSGVIFQRTLVLTLLDGTGARTGIRIOMEQSGAGSGLARYL			
	TVVWPLATQORSHOVNRPPLADILEOVLSPPAAYAAWRFAGAGDRMAFTRGSH			
	IAQRFRTVRFVTRLLIALAGLFTTVDEQASGHALITADSPRLAEDTASAGEG			
	RTHASQDERDAIQAICSRIVAGVVAAYAMDQAQKAPRAVNSRFPCGIASPDPA			
	YLSISPLADPTAAQRITAOQWESVVARALFGRGSVTRLSGTSLKTVCHPLP			

[illegible]

	gene	6745. .7527 /gene="RSP0178" /note="synonym: RS04690"
CDS	6745. .7527 /gene="RSP0178" /function="miscellaneous; unknown" /note="Product confidence : hypothetical Gene name confidence : hypothetical predicted by Codon_usage predicted by Framed" /codon_start=1 /evidence=not_experimental /transl_table=11 /product="HYPOTHETICAL PROTEIN" /protein_id="CAD17329.1" /db_xref="GI:17430646" /db_xref="GOA:Q8XTD5" /db_xref="TrEMBL:Q8XTD5" /translation="MTQYLHIITTTAVARIIRFKGLTLAAHPALGRPVARRHAEEV NRBAEGBGVNRLKAYIKKGLEAGYSLDIIRGDPFPIPVAPGNRDDEVEIIR VERAEVAFPLAAGCTPANKRGRTLMPLRTLGEHADMLPRKANALCRLAHVHSLEY AIEBGMTSRIVTFSRPBRASDCYSSTRHGGAACSVALRVSMMAAAPLLDDPSDRRA VMTRKLRFQOIEIMPAISVDLFTLNADRAAANMPFLTWS" 7518. .8021 /gene="RSP0179" /note="synonym: RS04691"	
gene	7518. .8021 /gene="RSP0179" /note="synonym: RS04691"	
CDS	7518. .8021 /gene="RSP0179" /function="miscellaneous; hypothetical/partial homology" /note="Product confidence : hypothetical Gene name confidence : hypothetical predicted by Codon_usage predicted by Framed" /codon_start=1 /evidence=not_experimental /transl_table=11 /product="HYPOTHETICAL PROTEIN" /protein_id="CAD17330.1" /db_xref="GI:17430647" /db_xref="GOA:Q8XTD4" /db_xref="TrEMBL:Q8XTD4" /translation="MELNAPDRMGPTPMRVVGGAIALSALALAPALADARKWQ DEPARAIVSEAOPEVNSGTHIPVLIHADPAVVAQVSTDFVHRNAVIPGSKRII GKELRKVNAREIWIIGLOIPASGTRLPLQATMRDSAGVADLSFGALLAMTS EPRIVPH" complement(8066. .8488) /gene="RSP0180" /note="synonym: RS04692" complement(8066. .8488) /gene="RSP0180" /function="miscellaneous; hypothetical/global homology" /note="Product confidence : hypothetical Gene name confidence : hypothetical predicted by Codon_usage	
gene	complement(8066. .8488) /gene="RSP0180" /note="synonym: RS04692" complement(8066. .8488)	
CDS	/gene="RSP0180" /function="miscellaneous; hypothetical/global homology" /note="Product confidence : hypothetical Gene name confidence : hypothetical predicted by Codon_usage	
Query Match	41.9%; Score 741.4; DB 1; Length 189050; Best Local Similarity 64.2%; Pred. No. 3.e-167; Matches 1130; Conservative 0; Mismatches 626; Indels 3; Gaps 1;	
Oy	2 TGGCTTCACGAGGCGCTTCCTCCGCCGGGTGAATTCTTGAGCGCTCGAGCAT 61	
Db	66492 TGGCTTCGCTGAGCCGCATCGTGCCTATCCCAAGCGGATGACGATGCTGTCGCG 6655	
Oy	62 GGGCTAAGACCCTGCAGAACAAACTGCTTCGCTCAGAGGCGCGCAATAGGAGAAATGC 121	
Db	66552 GGGCGGACAGCATCCGGATCGGCTCCTCGCGCGCAGCGCGCGCGCATGGCAGTGGG 6661	
Oy	122 GTGCTATACGTAAGCGGGAATGTTCCACAACGTCGCGCGGCATTCGACAGAAGCTTGTT 181	
Db	66612 AGGCAATACACTATGCGCACATCTGCACCGGCGCGCGCGGCTGCGCAGGCGCTGCG 6667	
Oy	182 CTTAAGCAATATGCGGCGAGAGCTGCGCGCTTATTCGCTCTGGAATAATGACCTGGAATC 241	

Db 6672 CGGAGGAGCTGTGCGCCGAGCGGCTTGCTGATCTGTGGGCAAGATCTGAGCATC 66731
 Qy 242 TTCAAGCTGGATTTGGGGCTATATATATGCGGGCATTTCCCTATTTGCCGCTCTCTT 301
 Db 66732 TGAAGCTGGGCTGGGGGCGCATATATGCGGCAATTCCTATTTGCGGCTGCGCGCT 66791
 Qy 302 ATTCACTGCTGTGCAAGATTTGGCGAAGCTGCTCATCTGTAGTCTTCTTGCAACGG 361
 Db 66792 ACCGCTGTGTGACGAGAGCTACAGCGGCTCTGCTACTGTGTGCGGACCTGACGCTG 66851
 Qy 362 GACTGTCTTTGCTGCGGATGACAGCATTTTCAGCGCGCAATTGAGACCAATTTGCGG 421
 Db 66852 GCTGTGTATGCGGCGAGCGGTGCTGTTCGCGGCGCGATCCAGGCGGTGTGCGCG 66911
 Qy 422 ACCAGTGTGCGGCAATTTCACTGTAGAGCGCAATTTGGCGGCGCGGCGAGCTGATTG 481
 Db 66912 CGAGAGCGAGGTGCGCATGACACAGCGAGGTGGGCGGCGCGGTGACCGGCTGG 66971
 Qy 482 ACAGCTGTGAGAGCAGCTGTGGAGTTGAGCAGATTAATGCTTTGCGGCACTGAGC 541
 Db 66972 CGTGTGCTGCTG--CGACGGGCGGCGGCGAGCTGTGATGCGGCGCAAGCGCGGTG 67028
 Qy 542 CCGATAGATTTGCGCAAGTCTTTGTTCACTTGTGCTTACCAACTGCTTAAGCGGTG 601
 Db 67029 CCGACAGATCGCCAAATTTCTGTGACTCTGAGGCTGACGCGGAGCCCAAGCGGTG 67088
 Qy 602 CGACTACTGAGCGCAATCTCTGCGCCCAATAGCAGATGCTTTTCCAAATTTCCGCTT 661
 Db 67089 CCAACAGCATCGCATCTGTGACAGCAACAGCAGATGCTGTGCAACCTTTCCCTGCT 67148
 Qy 662 TTGATGAGAGCGCGGCTGTGTGAGATGTTGCGCGGAAACACACTTTGCGGCGCA 721
 Db 67149 TCGGCGAGGCGCGGCTGTGTGAGATGTTGCGCGGAAACACACTTTGCGGCGCA 67208
 Qy 722 GCCACAAACATCGGATGCTGTGTGACAAAGCGGCGAGCTACTGACGAGCGTAAAC 781
 Db 67209 GCCACAAACATGAGGATGCTGTGTGACAAAGCGGCGAGCTACTGACGAGCGTAAAC 67268
 Qy 782 CAACCGGCGCAAGGTTTGGCGGAGCGCTTGGCACTTGAAGGAATCTTCTCCACTGCT 841
 Db 67269 CGACACCGAGGCTGTGAGACAGCGCTGCGCAACTGCGGAGTGCAGCGCGCTGT 67328
 Qy 842 ACCTCACTGTGCGAAGGTTTGGCGGAGGATTAATGAGGCTGCTTGGAGAGAGCATGCC 901
 Db 67329 ACTTCAACGATGCGGCGGAGGCTGTGAGAGTGTGACGAGCGCTGAGGCGGAGCG 67388
 Qy 902 TGGCGGAAAGCTTCTTGTGCTGCAATGAAGCTTCTTCTTGTGCGGCGGCTGTGCTG 961
 Db 67389 TGGCGGAGCACTTCTATGCGGCGGCTGCGGCTTCTTCTTGTGCGGCGGCTGTGCTG 67448
 Qy 962 AAGGATCTGTGATGCTTTGGAACGGGCTGCTGAACAGCATGTGTGAGCGCATTTGCA 1021
 Db 67449 AGGCGCGGTGAGACCGGCTGAGGCGGCTGACCGAGGCGCACTGCGGAGGCGATCGCA 67508
 Qy 1022 TGATGAGGCGGCTGTGAGCATGAGAGCTGCTCTTCTGCACTTTTACCAACGAGCGC 1081
 Db 67509 TCAATGCGGCGCTGTGAGCATGAGAGCGGCGGCTGTGCACTTTTACCAACGAGCGC 67568
 Qy 1082 TGTGATGCTGTGATTAATTTGAGCTGCGCAAGCGCTGTGAGAGTCAAGCTCTGCTCG 1141
 Db 67569 TCAATGATGCGGCTGTGATTAATTTGAGCTGCGCAAGCGCTGTGAGAGTCAAGCTCTG 67628
 Qy 1142 TCGATGAGGAATTTGGAAGGCGCTTTCATGTGCGGACGTCATAGAGGCGTACTGCGGTG 1201
 Db 67629 TCGAGCGCAAGCTGGAAGCGGCTTCAAGGCGCGCATGTCTATCGGCGTACTGCGGCG 67688
 Qy 1202 CTCTGGAACAAATTTGCGCAAGCGCTTTCATGAGAGGAGTAACTTGTCTCGGCTGATGCA 1261
 Db 67689 CGCGGAGCTGTGCGGCTGGAACGTTTCAAGAGAGCGGCTACTACGCTCGGCGCAAGCG 67748
 Qy 1262 TCAATTTGAGAGATCTGTGCGGATCTCAGAAAGGTGTGATGTTGAGCGGTGCAATTTGCTG 1321
 Db 67749 TGGCGCTTGTGAGAGGCGGCGCGGAGATCGGCTGTATGTTGACGCGCGCATCGCG 67808

Qy 1322 AAGACTTCAAGCTGTCTCAGGGGATTTTTCAGCGGCTTTGGGCAATGCGACGCGGCGG 1381
 Db 67809 AGGACTTCAAGCTGTCTCAGGGGATTTTTCAGCGGCTTTGGGCAATGCGACGCGGCGG 67868
 Qy 1382 TTTCGAGAGCGGCTTTTTCAGGCTGTGAGCGTGTGCTGTCTGATCTGATGATGCC 1441
 Db 67869 TCAAGAGAGGCGGCGCTTACGTGTGAGCGGCTGTGAGCGGCTGTGAGCGGCTGTGAG 67928
 Qy 1442 TTGATGCTGTGCTTTTTCAGGCTGTGAGCGGCTGTGAGCGGCTGTGAGCGGCTGTGAG 1501
 Db 67929 TCAAGCTGTGCTTTTTCAGGCTGTGAGCGGCTGTGAGCGGCTGTGAGCGGCTGTGAG 67988
 Qy 1502 AGGCTGTGAGCGGCTGTGAGCGGCTGTGAGCGGCTGTGAGCGGCTGTGAGCGGCTGTGAG 1561
 Db 67989 GCGCTGTGAGCGGCTGTGAGCGGCTGTGAGCGGCTGTGAGCGGCTGTGAGCGGCTGTGAG 68048
 Qy 1562 TCAACGACTTCAATCGAAGAGCACTGCGCAATGCGCATCATGATGAGGCTGCTC 1621
 Db 68049 TCGAGGCGCTGAAACCGGCGGCGGCGGCTGCGGCGGCGGCGGCTGCGGCGGCTGCG 68108
 Qy 1622 TCGATACGCGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1681
 Db 68109 TGGAGAGGCGGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 68168
 Qy 1682 GCGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1741
 Db 68169 GCGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 68228
 Qy 1742 CCATGCTGCTGTGAGAGGC 1760
 Db 68229 AGGTGATGCTGTGCGGAGCC 68247

RESULT 9
 BX572598 346879 bp DNA linear BCT 20-APR-2004
 LOCUS Rhodopseudomonas palustris CGA009 complete genome; segment 6/16.
 DEFINITION
 ACCESSION BX572598 BX571963
 VERSION BX572598.1 GI:39648490
 KEYWORDS complete genomes.
 SOURCE Rhodopseudomonas palustris CGA009
 ORGANISM Rhodopseudomonas palustris CGA009
 Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 Bradyrhizobiaceae; Rhodopseudomonas.
 REFERENCE
 AUTHORS 1 (bases 1 to 346879)
 Lartimer, F.W., Chain, P., Hauser, L., Lamerdin, J., Malfatti, S., Do, L.,
 Land, M.L., Pelletier, D.A., Beatty, J.T., Lang, A.S., Tabita, F.R.,
 Gibson, J.L., Hanson, T.E., Bobst, C., Torres, J.L., Peres, C.,
 Harrison, F.H., Gibson, J. and Harwood, C.S.
 TITLE Complete genome sequence of the metabolically versatile
 photosynthetic bacterium Rhodopseudomonas palustris
 Nat. Biotechnol. 22 (1), 55-61 (2004)
 JOURNAL 14704707
 PUBMED 2 (bases 1 to 346879)
 Lartimer, F.W. and Harwood, C.S.
 REFERENCES Rhodopseudomonas genome consortium
 AUTHORS Rhodopseudomonas genome consortium
 CONSRM Direct Submission
 JOURNAL Submitted (24-JUL-2003) Submitted on behalf of the Rhodopseudomonas
 genome consortium, the DOE Joint Genome Institute, Production
 Genomics Facility, 2800 Mitchell Drive, Walnut Creek, CA 94598,
 USA, and the Genome Analysis Group, Oak Ridge National Laboratory,
 1060 Commerce Park Drive, Oak Ridge, TN 37831, USA;
 lartimer@ornl.gov

FEATURES
 source
 1. 346879
 /organism="Rhodopseudomonas palustris CGA009"
 /mol_type="genomic DNA"
 /strain="CGA009"
 /db_xref="taxon:258594"
 /db_xref="taxon:258594"
 /gene="dapA1"
 /locus_tag="RPA1571"

CDS
127. .1005
/gene="dapA1"
/locus_tag="RPA1571"
/EC_number="4.2.1.52"
/function="InterPro IPR002220 COGS COG0329"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative dihydrodipicolinate synthetase"
/protein_id="CAE27012.1"
/db_xref="GI:39648491"
/translation="MMADFHGVPPYIVSPVDAGRYRADVMGRLCDLLIOAGVHGLTP
LDSGEFALVGTAGREAVVATIEAQRVPVAVGASVADVAQAALYELGLADG
IIAIEAYEPFLKDAQIESYPRADAVEIPVYITNPQFQRSLTLDVIALEHPR
RIKADSTWGRLLSIINRCGDALOVFSASAHIPAAVMILIGVMAGPACIAPROSV
ALVELCKAQRWDEALMLQKRLMRVNEAPAFNLAACIKGMLAQVDGDPPIPPQAL
TAEERKAVKVLAEIA"
1230. .3560
/locus_tag="RPA1572"
1230. .3560
/locus_tag="RPA1572"
/function="InterPro IPR000674 COGS COG1529"
/note="Observed by proteomics
Citation: Proteomics from Verberkmoes et al. (2003)
unpublished"
/evidence=not_experimental
/codon_start=1
/transl_table=11
/product="putative carbon-monoxide dehydrogenase large
subunit"
/protein_id="CAE27013.1"
/db_xref="GI:39648492"
/translation="NNILPGSMKFGAQGVKRLLEDORLLTGHGVLDDKAEGALMLV
VMSRPAHDAIVSIDTEAKGMGVPAVLGTADLIADVGITPLPIFKRPSPMAL
PIRLIAHLKRPVGEPPVAAYVATSOIAQAAMEAIAVYHEIPATDPPRAIQAP
ATVETPNDIVAMSYGDAKXDAFASAAHTSIDLVSGRLFSAMERSTIAIEK
KGRLLIHTQSTPQGTQDALAEAILIKRPRESIQVIVGIDGGFGQKTGVYPPDALVA
YAAVKLRIRKIRMGERTDEFGTGRDLTSTFALDSKGLAVRVISIGGTGYL
AGAGVVIPLVAGPQVGTGVLPHVAFDIKAVWTHPAVGAAYAGAPKAVVYVERLM
DAAARKMLDPRAIRKINYIKPSQLPTNAVGVGSAFAHMDRAELSDGDFKA
RKKAQKGLVGRGVTSYIEWGSRPHNTYHATGARGVILHSGTAMGGGLT
YTMIAALEIFLDKIDVQGNTDLAIGFSVSRSLLPVGTRVAVASAMTAKAREK
AAHTEASIEDIEYSGMLTIAGTDRKISLFTIAAEKGRSLSDSTGEVDGSPWNG
AHICEVVDPEGTESRVVRYTVDVGNVPMVLVAGIHHGVAQGVQALYENAAVN
DGGQLTASYLDYICIPRADNLPIINVLDPSPACKTNPLGAKGCGESGAIGPFCVH
GVLDALAPLGVTTLNTPTEPKVRAIQAKSAQA"
3638. .4252
/locus_tag="RPA1573"
3638. .4252
/locus_tag="RPA1573"
/function="COGS COG1704"
/note="Observed by proteomics
Citation: Proteomics from Verberkmoes et al. (2003)
unpublished"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="LemA family"
/protein_id="CAE27014.1"
/db_xref="GI:39648493"
/translation="MRKFLTVLAAALALSLNSGNYTTIOSEDEVKSTWSEVVNOYR
RADLVNLVNSVGFQAQBERDVLGVTNARAKGSIQATPEVNLNDPAQKPOQAGE
LSASLSRLIIVTENYFQPKSDLEFKMLMAQLBGTENRITVARNRVTKAAVAAYVTVS
VPIPTAMMGVYKKNFTYENKAIISTAPKVDFTAPAPAPAK"
3638. .3700
/locus_tag="RPA1573"
/note="Signal predicted by SignalP 2.0 HMM (Signal
peptide probability 0.998) with cleavage site probability
0.396 at residue 21"
4257. .5180
/locus_tag="RPA1574"
4257. .5180
/locus_tag="RPA1574"
/locus_tag="RPA1574"

misc_feature
sig_peptide
sig_peptide
/function="COGS COG1512"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="DUF477"
/protein_id="CAE27015.1"
/db_xref="GI:39648494"
/translation="MTAAITRAVRAGMALLCVVALAPAPADAVAPVLTGRVVD
QGTTSADAVARDQKLKAFEARKGSQIVLVPTTPQPAIEQFSIRVAEMKIGRK
IDDGATILVAKDNRKLRIEVGLGALPDVAKRIIDETIPKFTGPDGSIKAGV
DMIGVIGDEPILPTPOKHEMAAPESMDLTGMATPLDGLTINQNLKALAGRVAS
GTTGTVGVANPLFGLVWITALVAGFAATLIVLDLFGPTTSSGRSGSGTWTG
SSGSWSWSSSSSGFGGSGSGFGGASGSW"
4257. .4352
/locus_tag="RPA1574"
/note="Signal predicted by SignalP 2.0 HMM (Signal
peptide probability 1.000) with cleavage site probability
0.947 at residue 32"
order(4290. .4358,4824. .4880,4899. .4952,4962. .5030)
/locus_tag="RPA1574"
/note="4 probable transmembrane helices predicted
by TMHMM2.0"
5194. .5691
/locus_tag="RPA1575"
5194. .5691
/locus_tag="RPA1575"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAE27016.1"
/db_xref="GI:39648495"
/translation="MGIRIGKHLPTSRGAVHRAFSPPDAFDAIERAIKASETHVQGI
RFVREGALGALPFRDQPVBERALDFSLRLIMDTFNNNGVILYLALARNVYIIDR
GIDRGAGTWTAEIACREMAEFREGEFEGVLRGIDRIHAHLAEHPRSGSPNEIPD
APVVV"
complement(5773. .6384)
/locus_tag="RPA1576"
complement(5773. .6384)
/locus_tag="RPA1576"
/function="InterPro IPR004045: IPR004046 COGS COG0625"
/note="Observed by proteomics
Citation: Proteomics from Verberkmoes et al. (2003)
unpublished"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative glutathione S-transferase"
/protein_id="CAE27017.1"
/db_xref="GI:39648496"
/translation="MKLYDSIGPNPRLVPMFTAEKGLITPTQTLDIRAGENRKEPIHA
RPHGOTPTLELDNGSFLEAVTAICYLEEKHQPPLIGTPEERAECPMWTIRIDLY
ICEPLNGRFAEGK.KMFQDRVVCVPEAAGLKXIAADRIIMNNEGMKGEYCGARL
TIADLLIYGMIDPGAOVQPLAPANTTIEMFRITARSAA"
6593. .7564
/locus_tag="RPA1577"
6593. .7564
/locus_tag="RPA1577"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="CAE27018.1"
/db_xref="GI:39648497"
/translation="MDEHNEPIRGEOGTAAATGGAONAGTPEASSASAPOGTTPAN
TPANAGAKPDEAKPRAGKSEIKSDRSVPFRGTYTTPASPRREDIYRIMWDEIYVEA
PDBASQSAVHRGVRKLAIVAIVIAIAGVAGLATAISHSRSDATTAASAVYK
TLDQVAGQOIEISALKADLDRAKQSTQIGKTDRIKLEKQAEPMVAKTAKLSY
VEKLRTAQVPPAAPANVAAPKADVYGSITPKKAAKPGRLPTVEGVLADVVDG
GAVBGRSGTYEVYAGDPIGVGRVDAIRQDGRWVWVTSRGLIVAR"
6947. .7015
/locus_tag="RPA1577"
/note="1 probable transmembrane helix predicted


```

terminator      7593..7621
                  /note="TransTerm prediction, 98+%"
gene            complement(7721..8494)
                  /gene="fpr"
                  /locus_tag="RPA1578"
                  /locus="RPA1578"
                  /locus_tag="RPA1578"
                  /locus="RPA1578"
                  /function="InterPro IPR001433:IPR001834 COG6 COG1018"
                  /evidence=not experimental
                  /codon_start=1
                  /transl_table=1
                  /product="ferredoxin-NADP+ reductase"
                  /protein_id="CAE27019.1"
                  /db_xref="GI:39648498"
CDS
Query Match      29.8% Score 527.8; DB 1; Length 346879;
Best Local Similarity 56.6%; Pred. No. 7.7e-116;
Matches 976; Conservative 0; Mismatches 747; Indels 0; Gaps 0;

2   TGGGTTCTCGAGGCGCTTCTTCCCTTCCGCGGTGCAATTCTTGAGCGTCTCGAGATT 61
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  256342 TGGGCTGACACACACGCTCACCGACTATCGGTGCGGATTACGACCGGCTGCATCACT 256401

QY  62  GGGCTAAGACCCGTCAGAAACAACCTGCGTGTCTGCCAGGCGGCAATGGGAATGCG 121
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  256402 TCGCGAGACGCGCCGACCGGCGTTCATGCGCGAGCGMAAGTGAGCGCGCTGCG 256461

QY  122 GTGGTACAGCTACGGGAAATGTTCAACAACGTCGCGGCATGACAGAGCTTGCTC 181
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  256462 GCAAGATCAAGTATGCGAGATGCTGCGCGCGGACAGACCATGCGCTGCGCTATCG 256521

QY  182  CTTACGACTATCGGACAGCGCTCGGCTTATCTGCTCTGAAATGACCTGGAATC 241
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  256522 CGGCGGACGTGCGCGGAAACGCGCGGTGATTTGTCGGTAAATTCATGACATG 256581

QY  242  TTCAAGCTGGCATTTGGGGCTATGATGCGGCAATTCCTATGCGCGTCTCTGCTT 301
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  256582 CGATGATGATGTTGCGGCGCTGATGCGGCGCTGCGATGTCGGTGTGCGCGCGT 256641

QY  302  ATTCACTGCTGCGAAGATTGGCGAAGTGGCGTCACTGTAGTCTTGCAACCGG 361
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  256642 ATTGCTGCTGCTCAAGGATTTACGCAAGCTGCGCATATGTCGGCTGCTGACTCGG 256701

QY  362  GACTGATCTTTGCTGCGGATGACACACTTTCACGCGGCAATTGAGCACTTTCGCGG 421
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  256702 GGGTGAATCTTCGCGGATGACATCACTGCTTCGCGCGCGATCTCGGCAACGCTGCGG 256761

QY  422  ACGACGTGCGCGCAATCTTCACTGAGCGGAATTGGCGGCGGCGGCAAGTGAATTTG 481
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  256762 AGGATGTCAGAGCTGCTGCCACGCGCGGCGAGTGAAGGGGCGGAAGTGAAGTGCCTTG 256821

QY  482  ACAGCGCTGGAGACGCTGCTGGGAATTGAGCGAGATTAATGCTTTGGCGGCACTGGCC 541
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  256822 CCGAATGCTGCGGACGCGGGAACATCCGAACTGCGCGGCAAGACGAAACGATTCGCGC 256881

QY  542  CCGATACGATTCGCAAGTCTTCTTCTTCACTTCTGCTTACCAAACTGCTTAAGCGCGTGC 601
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  256882 ACGACACATCGCTTAAGTTTGTGTAAGTGGGATGACCGGCAATTCGAAGCGGGA 256941

QY  602  CGACTACTCAGCGAATGCTTGGCGGCAATCAGCAAGATGCTTTCGCAAACTTTCGCGGTTT 661
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  256942 TCAATACGACGCGGATGATCTGCCCAATCAAGTGAATGATCCGAGAGCGATGCGCTTCC 257001

QY  662  TTGGTGAAGAGCGCGGCTGTGTTGATGATGCGGTGGAACACACCTTTCGCGGCA 721
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  257002 TGAAGAAGACGCCCGGTGATGCTCACTGCTGCGGTGAATTCACCTTTCGCGGCA 257061

QY  722  GCCACAACATCGGATCGTGTGTTGACAACGCGCGGCACTACTTACCTTGACGACGATTAAC 781
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  257062 ACCACAATATCGGTCTGACGCTGTTCAACGCGCGGCTGATGTATATGACGACGCGCAAGC 257121
```

```

QY  782  CAACCGCCAAAGGTTGCCGAGACGCTTGGCAACTTGAAGGAATCTTCCCACTGCGT 841
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  257122 CGAGCGCGCGCGGATGCTTCCACATCCGCAATCTCGGGAATGCGCGACGGTGT 257181

QY  842  ACCTCACTGTGCCGAAAGGCTGGAGGAATTAATGAGTGGCTTTCAGAGACACATACCC 901
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  257182 ATTTCAACGTTCCGAAGGCTTACGAGTGGCTGCTGCGCGGTCTGCGGGAACACAGCAAGT 257241

QY  902  TGGCGAAGCCTTCTTCTGCTTCGATGAGGCTTTCTTTCGCGGCGGCTGGGTTGTCGC 961
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  257242 TGGCAATATGTTCTTCAAGCGGCTGCATGATGTTCTTCTCGGCGCCAGCTTCGCGG 257301

QY  962  AAGGATCTGGGATCTTTTGAACCGGCTGCTGAACAGCACTGTGTGAAGGCAATTGCGCA 1021
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  257302 CGCATGCTTGAAGAGGCTTGAAGAGTGTGGGTGCGGAGAACCGGCGCGGATGCGGA 257361

QY  1022 TGATGCGGCTGCTGGGATGACGAGACTGCTCTTCTGCTCACTTTTACACCGGACCGC 1081
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  257362 TGCTCACCGGCTTGGGCGCACGAGACCGCGCTTCTTCAATGTCGGTACCGCGGAGA 257421

QY  1082 TGTGATGCTGTGTTACATTGGGCTGCGCAAGCGCTGCGTGGAGTCAAGCTGTTCCGG 1141
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  257422 CCAATGCTCCGCGCATGTCGCGCTGCGCGGTGCGGCAACGAGCGCAAGCTGTGCGGA 257481

QY  1142 TCGATGGGAATTGGAAAGGCGTTTCATGATGTCGCAACATATGAGGGCTTACGCGGTG 1201
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  257482 ACAACGGCAAGCTGGAAGTCCGCGCAAGGCGCGGAACATCAACCCGCGATTGGGCGG 257541

QY  1202 CTCCTGAACAAATGCGCAAGCGTTGACAGAGAGGAAGCTTATGCTTCCGGTATGCCA 1261
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  257542 CGCCGAGCTGACCGGATTAAGCGCTTTCAGACAGAGGGCTTCAACACTCAAGATGCGC 257601

QY  1262 TCAAAATGGCAATCTGCGGATCTCAAAAAGSTCTGATGTTTGAAGTTCGAATTCGTG 1321
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  257602 TGAAGCGGTGATGCAACAGACCTTTCGCGCGGCTTCAATTCGAGCGCGGATTCGCG 257661

QY  1322 AAGACTTCAAGCTGCTCTGAGGGGTATTGTACGCGTTGGGCGCATGGCAACGCGGCGG 1381
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  257662 AAGACTTCAAACTGGCTGCGGACCTGGGTGAGCGTGGTCCGCTACGCGCAAGTTCA 257721

QY  1382 TTCTGAAAGCGGCTCTTACGCTCTGAGACGATGATGTTGCTGCTCTGATGTAATGCC 1441
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  257722 TTGCGGCTGCGCTCTCCCTGCTGGTGGCGACGTGTGATGCGCGGCTGACCGGATTAAG 257781

QY  1442 TTGATTTCTGTGTTTCCGCGTCTTCTCACTGCGCTGCTTGTGCGGCTTGAAGAAAG 1501
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  257782 TCACCGGCTGGCGATCTCGATCCGACGCGCTGCAAGCTGATCAATGCGACGCTGCGCG 257841

QY  1502 AGGCGTGGACGCGGAGTGTCTGCGCACTGAGCCGCTTTCGGGCTGTGCTGATGCGG 1561
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  257842 TGAAGAAGCTTCGCGGATGCGCGGCGGCAACATCTGATTCGGAGGCGCTTCCGCAAGCGCT 257901

QY  1562 TCAACGACTCAATCGAAGAACAACTGGCAATGCGCAATGCGCATGATGAGTGAAGGCTCC 1621
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  257902 TCGGCAAGCTGTGACGCAAGGCAACGCGCTGCTGCAACCGGCTGACCCGCGCGGTGCTGC 257961

QY  1622 TCGATACGCGCGCTGATGATTAAGGCGAGAGTCACTGACAGAGGCTTGATCAACAGC 1681
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  257962 TCGGCGAACCGGTGCTGATGACAAAGGGGAGATCAACGCAAGAGGCTCGGTCAACAGC 258021

QY  1682 GCGGCTTTTTCGAATGGCGGCTGCGGAAAGTTGATGCGGCTGTA 1724
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  258022 GGGCGGTGTAATATGCGCGCTGTTGATGCGGATCTTTA 258064
```

```

RESULT 10
AP005963      299950 bp   DNA      linear   BCT 12-SEP-2003
LOCUS        Bradyrhizobium japonicum USDA 110 DNA, complete genome, section
DEFINITION   29/31.
ACCESSION    AP005963
VERSION      AP005963.1 GI:27355938
```

[illegible]

comp1ement (4162. . 4824)
 /gene="b117663"
 /note="ORF ID:b117663
 unknown protein"
 /codon_start=1
 /transl_table=1
 /protein_id="BAC52928.1"
 /db_xref="GI:27355947"
 /translation="MKRMRYASISSTVIAVGVGACALVVGOSHPVDAIASSVTRAP
 ELMEARLIPVATFOFHVDQMSNGSCGPAVAVNARYELGEARTREGVLGWSCW
 TGVCITMGLITDELIAVACQNTSRKTYVLDLSENQLEHLRRANDRGRIYIVNPDAR
 IFQASGHSHP1GGTFFEADLVFVLVDNVPFQWVERKRLFDAVNTLDGDKRGILL
 IE"
 complement (7319. . 9061)
 /gene="b117664"
 complement (7319. . 9061)
 /gene="b117664"
 /note="ORF ID:b117664
 putative adenylate cyclase 3 (EC 4.6.1.1)"
 /codon_start=1
 /transl_table=1
 /protein_id="BAC52929.1"
 /db_xref="GI:27355948"
 /translation="MTEQVQRRLAIIADLVGSGALMQRABEATYAEPERIKRELI
 EPSLRHSDGRILKKTGDGALAPFAPPLAVRATQEMDILAGSGEPFKIRIGMLDVI
 IVGGDGLYGDGINIAVRLEGIADPGILMSKRVSEVGKLDVGEDDEGQOLKIIA
 KPVRAVYRAGASALSALTERLSAAPLPAPKPSIAVLPFENMSGDPEDYEADGVVEIT
 TALSFEKMLFVILARNSSFTYKKAIVNKGVHELGVRYTLEGSVYRAAKRVYIAQGLI
 DAITSTHIMADRFERDLDIPALQDDVALAVSADVPKMLQTEIATNARRPENLTYA
 DPEFLSKQOYVYITRESLEAEIKILAHRALELDIPRESSAALAGDCHASVVALGYANP
 EFDNAEAIRLRLALSLDDGDDPTLSRVAVSVACVGDSEAEIEMADRAVVALPNPSWR
 TWNNQGVYTMAGLPEBAIRSPERAIIRMSPLDPLMLMTFLGIALVGRFPEALIV
 LKALRONSFPWPCACGLASAPAHIGHGRDAEVAHAARVALEIDPTFTSGHIAVRPSA
 KMFLEGRKACGPE"
 9440. . 10444
 /gene="b1x7665"
 9440. . 10444
 /gene="b1x7665"
 /note="ORF ID:b1x7665
 putative methyltransferase (EC 2.1.1.-)"
 /codon_start=1
 /transl_table=1
 /protein_id="BAC52930.1"
 /db_xref="GI:27355949"
 /translation="MTSNPWLTLGLINGFOVTOAIVHASTLRIDHLKGDSSAEEL
 APLTQSHGALYRLRLAFAVSVLEDEDRRALTPMGSCATDSTPTPIGMAAECSGS
 PYWSTGWHLHLSIRTGENAFQSLGKDIWNTNRAERPEBSAFDRAMTEFSSRGAAV
 IGAUDFSAFHHVADVGGGGLMLAAILTDHPMPOGLTFQOPGVIAEAEVLEHAGVD
 RCRWSSGFFSEFPEGSDAYVVRVVIHDDDDDEAIIAIIKACRRAMRETAKLVIERII

QY	302	ATTCACTGCTGTCCGAAGATTTGGCGAGCTGCGTCACATCGTAGGCTCTTTTCGAACCGG	361
Db	199581	ATTTCCTGATGTGCGAAGATTACGGCAAGCTCTCTATGTGATGAAGCTCTTGACGCCCG	199640
QY	362	GACTGATCTTTGCTCTCCGATGCGAGCACTTTTCAGCGGCCAATTGAGACATCTCTCGG	421
Db	199641	GCCTGCTCTTGCCCGAAGAGACGCCACAAAGTTCTCGATGCGCTCGCTGCCAATGTTTCC	199700
QY	422	ACGAGCTGCCCGCAATCTTCACTCGAGGGAATGACCGGGCGCGCACGTGATGTTTTC	481
Db	199701	TGGGACCGAGATCCCGGCTCTACGGCCAGTGTGGCGGGCGCGACGTACGCTGCTCG	199760
QY	482	ACAGCTGCTGAGACAGCTGCTGGGATTGAGCGAGATTAATCCCTTTGGCGAATCGGCC	541
Db	199761	CCGACCTCATGGCGAGCGCCGATCCCGGCGACTCGACGAATCCACGCGAAGATGGCC	199820
QY	542	CCGATACGATTGCGCAAGTCTTGTTCATTCTTGAGCTTACCAACGTGCTTAAGCGCGATC	601
Db	199821	CCGATACGATTCGCAAAATTCGTGTGAGGTGGGGCTCGACCGGCAATCCCAAGGCCGTCA	199880
QY	602	CGACTACTCAGGAATGCTCTTGGCCCATCAGCAGATGCTTTTGCAACTTTTCCGGTTT	661
Db	199881	TCMACACCGAGCGATGATCTGCGCAACCAAGTATGTGCGCGAGACCTTCGCTTCC	199940
QY	662	TTGTGGAAGAGCCCGCGGTGTGTGTGAGACTGTGTGGCCGTGGAACCACTTTCGCGGCA	721
Db	199941	TCMAGAGAGAGCCCGCGCTCATCATGACTGTGCTGCCCTGGAATCACACTTTTGGCGCA	200000
QY	722	GCCACACATCGGCATCTGTTTGTACACAGCGCGGACGTACTACTTGAACGAGTAAC	781
Db	200001	ACCAAAATATCGGCTCTGACGCTCTCAATATGGCGGCTCGATGTATCTGACAGCGCGCAAC	200060
QY	782	CAACGCCCAAGGGTTCCGCGAGACGCTTCGACCTGAGCGAAATCTCTCCACATCGCGT	841
Db	200061	CGATGCCGATGCAATCGAGAGACCGTGCAGATCTCCAGAGATTTGCCGACCGTCT	200120
QY	842	ACCTCACTGTGCGGAAGGCTGTGGAGAAATTAGTGGGTGCCCTTGACGAGACAGTATCC	901
Db	200121	ATTTCAGTGTCCGAAGGGCTATGAGTGTGCTGTGCGCTCATCTGCGCGACGACAGAGGCC	200180
QY	902	TGCGGAAACGCTTCTTGTCTGCGATGAAAGCTTCTTCTTCCGCGCGCTGGTGTCCG	961
Db	200181	TGCGGCGAAATCTTTCGACCGGCTGCATATGCATATGTTCTTCTCGGGCGGCGCTGTCCG	200240
QY	962	AAGGATCTGGGATGTTTGTGACCGGGTGTGCTGAACAGCACTGTGTGAGCGCATTCGA	1021
Db	200241	CGTTGCTCTGGGACAGCTTCGACGAGCTGTGCGGTGAAGAAAAGCTATTCGCTGCCA	200300
QY	1022	TGATGGCGGCTGTGGGCATGACGAGACTGCTCTTCTCTGCACTTTTACACCGGACCG	1081
Db	200301	TGCTGACCGGCTTGTGGCGGACGAGACCGCGCCGCTTTCATATGTGCTCAATCCCGCA	200360
QY	1082	TGTGATGGCTGTATTACATTTGGGCTGCCAGCGCTGCTGGCGAGGTCAAGCTCGTTCCGG	1141
Db	200361	CCAGCGCTCGGCGCATGTGTGCGCTTCCGCTTCTCGGACACACGCAAGAGTGTGTCCCA	200420
QY	1142	TGATGAGAAATTTGAGAGGCGCTTTCATATGTCGCACTCATGAGCGGCTACTGCGCTG	1201
Db	200421	ACAAAGGCAAGCTCGAAGTCCGCGCCAGAGGGCGCAAGTATGCCCCGCTACTGCGGCC	200480
QY	1202	CTCTCGAACAAATATCCCAAGGCTTTCGACGAGGAAGGCTATTAATCTGCTCCGCTGATGCCA	1261
Db	200481	AGGCGGATATCCCGCAAAATCTTTCGACGAAGAAAGATTTCAAAAGTTGGGCGAGCGCC	200540
QY	1262	TCAAATTTGGAGATCTGCGCATCTCAAGAAAGTCTGATGTTTACCGTCAAAATGTCTG	1321
Db	200541	TCAAGCTTCGCGATCCGATATATCTCAACGCCGCGCTGATTTGACGCGCCCGCTACAGG	200600
QY	1322	AAAGCTTCAAGCTGTCTCAAGGGATTTTGTACGCGTTTGGGCAATTCGCAAGCGGCGG	1381
Db	200601	AGGACTTCAAGCTTCGCGAGCGGTACCTGGGTCAAGTGTGCGCCCTTTCGCGCGCGCTTGG	200660

QY 1382 TTCTGGAAGCGGCTTTACGCTCTGAGCACTAGTGTGCTCTGATCGTGAATGCC 1441
Db 200661 TTCCGGGCTCTCGACCGCTGTGGCGACGCTGTCATCGCCGGCATCAACCGGACGAGG 200720
QY 1442 TTGGATTCCTCGTGTCTTCCGGCTTCTGACTGCGCTGCTTGTGGGGCTAGAAAAG 1501
Db 200721 TCTCCGGCTGTGCTGTGCTGACCTCGACGCTGTAGGCTGTCMAACCGACGCTGCCG 200780
QY 1502 AGCGCTCGACGCGGAGGCTCTTCCAGTACGCGGCTTGGGCTTGTGCTGACTGCG 1561
Db 200781 CCGACGACCTCTGCTGTGCGACGCGGACCGGCTGTGGCGGAGCCTTTTCGAGAGGCC 200840
QY 1562 TCAAGCACTCAATCGAAGAACCACTGCAATGCACTGCACTGATGAGGCTCC 1621
Db 200841 TGACCGGCTTCTCGGCTCGGCGCACCGTTCTTCAACGCGGATCAACGCGCATCTGA 200900
QY 1622 TCGATAGCGCGGCTGATGATTAAGGCGAGGCTCACTGACAAAGGCTCGATCAACGAC 1681
Db 200901 TGAACACGCGCTCTCTCATGACAAAGGCGAGGTCAACGACGAGGCTCGATCAACGAC 200960
QY 1682 GCGCTGTTTCAATGCGGCTCGGCGAAGTTGATGCTGTATCGTG 1729
Db 200961 GCGCGTGTCTGAGACATCGGCGCGCTGATCAACGACTCTACGCTG 201008

RESULT 11
AR319843 1902 bp DNA linear PAT 17-AUG-2003
LOCUS AR319843 Sequence 2393 from patent US 6562958.
DEFINITION AR319843
ACCESSION AR319843
VERSION AR319843.1 GI:33700946
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1902)
AUTHORS Breton, G. and Bush, D.
TITLE Nucleic acid and amino acid sequences relating to *Acinetobacter baumannii* for diagnostics and therapeutics
JOURNAL Patent: US 6562958-A 2393 13-MAY-2003;
FEATURES
source location/Qualifiers
1..1902
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Query Match 25.2%; Score 445.4; DB 6; Length 1902;
Best Local Similarity 55.2%; Pred. No. 4,7e-96;
Matches 932; Conservative 0; Mismatches 746; Indels 9; Gaps 3;

QY 46 GAGCGTCTCGAGCATTTGGGCTAAGACCCGTCAGAACAAACCTGCTGTCGACGAGCG 105
Db 166 GATCGTTAATTCATTTTGGACAACCAACCGACATATTTTGGACGAAAACGCAAT 225
QY 106 GCAAAATGGGAATGGCGTCTGATCAGTACGCGGAATGTTCCACAGTCCGCGCATC 165
Db 226 GCTCAAGGCAATGGGTCAAACTGAGTTATGCAAGAACTTCTACAAACCGCATGGCAATT 285
QY 166 GCACAGAGCTTGTCTCCCTTACGAGCTATGCGAGAGGCTCCGCTTATGCTCTGGA 225
Db 286 GCTCAAGCTTTGATGCTGCTGATTTTAAGCCAAAGAACCTTGTGATTTTAAGTGT 345
QY 226 AATGACCTGGAACATCTTCACTGCGCATTTTGGGCTATGTAATGAGGACATTCCTATTCG 285
Db 346 AATGATCTGAACATTTTAACACTGTATGCTGATGCTGATGCTGAGGCGTCTTTCG 405
QY 286 CCGGTGCTCTCTGTTATTAATCACTGCTGTGCAAGATTTGGCAAGCTGCGTCACATCTGA 345
Db 406 GCTATTTCCCTGCTACTCTGATTTCTCAAGACTTTGGCAAACTCAAACTATGTGTT 465
QY 346 GGTCTTCTGAACCGGAGCTGTGCTTGTGCGGATGACACACTTTCACGCGGCAATT 405
Db 466 GAAGTCTCAACCTGTATGTATGATGACGATGACAACTTTTGGCAAGCGATT 525

QY 406 GAGACCATTCCTGCGGACGACGTCGCCGAATCTTCACTCGAGCGAATTGACCGGCGG 465
Db 526 CAGGCAATGATTAATGCGCTGATATTGAAGTGTCACCAATTAAGGAATATGCGGCGATCAG 585
QY 466 CCGACGCTGAGTTTGAACAGCTGCTGAGACAGCTGTGTGGATTTGAGGCAAGTAATGCC 525
Db 586 ATCTGCAAGCTTTTCAATCGCTGTAGATACCAAGTTTCAAAATGTTCAAGAGTTT--- 642
QY 526 TTTGCGGCAACTGCGCCCGATATAGTTGCAAGTTCTGTGTTCACTTCTGCGCTACCAAA 585
Db 643 TATCAAAACCTTATGAAAACAGATTGCCAAATTTCTGTTTACATCAGGTTCAACTAAA 702
QY 586 CTGCTAAGCGGCTGCGCACTACTCAACGAAATCTGTGCGCCAAATCAGCAGATGCTTCTG 645
Db 703 TTACTTAAGGCTGTACGACACACAAATTAATGTGTGTTAATCAGCAAAATGTTATTG 762
QY 646 CAAACTTTCCCGGTTTGTGTGAAGCCGCGGTGTGTGATGATGTTGCTGCGTGGAAAC 705
Db 763 CAGACTTTCCTGAGTTTGAAGAAACACCGCTGTCTACTCGACTGCGCTGTCTTGGCAC 822
QY 706 CACACCTGCGGCGGACGCCCAACATCGGATGCTGTGTAACAAGCGGACGTACTAC 765
Db 823 CACACATTTGGCGGACGTCAAAATGTGCGCATGCACTCTAATAACGCGGTACGATTTAC 882
QY 766 CTTGACGACGCTTAACCAACCCGCCAAGGTTGCCGAGACGCTTCGCAACTTGACGAA 825
Db 883 ATGATGATGAGCAAAACCGTTGACAGAAATTTGACGAACTATGTGTAATCTCAAGAA 942
QY 826 ATCTCTCCACCTGCGTACCTCACTGTGCGGAAGGCTGGAGGAATTAATGAGGTGCCCT 885
Db 943 ATTTCTCAACTGTATTATTAATGTGCCAAAGTTGGGAACACTCACCGAAGGTTA 1002
QY 886 GAGCGAGACGTAACCTGCGCGCAACGCTTCTGCTGCAATGAGGCTGTCTTCTGCGG 945
Db 1003 GAAAAATGATGAAGATTTAAGAGGCGCTTTTGGCCAAAGTTAAATTTATCTTTGCC 1062
QY 946 GCGGCTGAGTTGTGCAAGGAGATCTGAGATGTTTGGACCGGCTGCTGAACGACACTGT 1005
Db 1063 GGTGTGCACTTTCAAGAGGCGGCTGAAACAGACTGATTAATTTGCTCAGCAAACTTGC 1122
QY 1006 GGTGAGCGCATTTGGCATATGCGGCGTCTGGGATACCGAGACTGCTCTTCTGCACT 1065
Db 1123 GGAAGAAAAATCGCATTTATGAGCGGATGCGGATACCGAACTGCTCTTGTGCT 1182
QY 1066 TTTAACACCGGACCGCTGCTGATGCTGTGTTACATTTGAGGCTGCGACGCTGCTGCGAG 1125
Db 1183 TTTACAACTGCGCCACGCGTAATGCTGTGCTTTATTTGTTAACCTGCTCCGGATGCGAA 1242
QY 1126 GTCAAGCTGCTTCCGCTGATGAGGAAATTGAAAGGCGTTTCCATGCTCCGACGTCATG 1185
Db 1243 ATTAAGCTAGTTTCATGTGCTGACAACTTGATTTTGCCTGTGTGCGCAAAACGTCATG 1302
QY 1186 AGCGGCTACTGGG---TGTCTGTGAACAAAAGGCCCAAGCGTTGACAGGAAAGGCTAT 1242
Db 1303 AAAGGCTATTGGGCTTAAAGGCGACCAACAAAGTACTATTTTATGATGTAAGGCTTT 1362
QY 1243 TACTGCTCGGATGATGCAATTAATGGCAGATCTGCGCATCTCGAAGAGGCTGTGATG 1302
Db 1363 TTCCATACAGGCGATGCGCTGTTAATGTCATGATGATCAACAAAGGCTTAATG 1422
QY 1303 TTTGACGCTGGAATTTGCTGAAGCTTCAAGCTGCTCAAGGAGTATTTGTCAAGCTTGGG 1362
Db 1423 TACGAGGAGCAATTTCCGAAGACTTTAACTCAATACAGGCACTTTTGTGAATGTCGGG 1482
QY 1363 CCATTTGCGACGCGGCGGCTTGTGAAGCGGCTCTTAAGTCTCTGACACGATGATGCTGT 1422
Db 1483 ACACTACGGAACAAATGCTTATTTCAAGGTAATTAATCAATCAAGATGTTGTATTAAT 1542
QY 1423 GGTCTGATGATGAATGCTTGAATGCTGTGTTTCCGCGCTTCTGCACTGCGGCTC 1482
Db 1543 GGTTCAAACCTGAATGCTATTTGTTTCTGATTTTTCAAAAATTAAGACGCTTGTGCTCAA 1602

QY	1483	TTGTGGGGCTGAGAAAGAGCG---TCGACGCCGAGAGTGCTTCCAGTGAGCCGGTT	1539	MEDLINE	94341565	
Db	1603	TATGAGGCTTTTAAAGCTGGCGCAATATTTCTTGACAGAGATATCTCAACACCTTAAAGTTC	1662	REFERENCE	8063101	
QY	1540	CGGGCGCTGTTTCTGACTGGCTCAACGACATCGATCGAAGCAACTGGCAATGCCAGT	1599	AUTHORS	6	(bases 1 to 45324)
Db	1663	CACACATGTTTCCGCCAATTTTAAACGACTTTTAAAGATGCGACTGGCAGCTCAAT	1722	TITLE	Elemore,D.A. and Ornstom,L.N.	The pca-pob supraoperonic cluster of <i>Acinetobacter calcoaceticus</i> dehydrogenase
QY	1600	CGCATCATGTGGTGAAGGCTCCTCGATACGCCGCCGTGATGATGAAGGGCGAGTCACT	1659	JOURNAL	J. Bacteriol.	176 (24), 7659-7666 (1994)
Db	1723	ACAGTCTCAATCTTTATTTAATGACGAGCCACCTCAGTTAGTCCGGCGAAGTGACC	1782	MEDLINE	95095936	
QY	1660	GACAAAGGCTCGATACACGAGCGCTGTTTTCATAGCGCGTCCGCCAAGTTATGCG	1719	REFERENCE	7	(bases 23320 to 27131)
Db	1783	GATAAAGCAACCTCAATCAAGCAGATTACCAAGCGTGGCGCTTTATTTATGATGAG	1842	AUTHORS	Elemore,D.A. and Ornstom,L.N.	Unusual ancestry of dehydratases associated with glutamate catabolism in <i>Acinetobacter calcoaceticus</i>
QY	1720	CTGTATC 1726		JOURNAL	J. Bacteriol.	177 (20), 5971-5978 (1995)
Db	1843	CTTTATC 1849		MEDLINE	96011389	
RESULT 12				AUTHORS	7592351	
ACCPAOP/c				REFERENCE	8	(bases 1 to 45324)
LOCUS	ACCRAP	45324 bp	DNA	linear	Gerischer,U., Segura,A. and Ornstom,L.N.	pcaU, a transcriptional activator of genes for protocatechuate utilization in <i>Acinetobacter</i>
DEFINITION	Acinetobacter sp. ADP1	dca-pca-qui-pob-hca	supraoperonic gene		J. Bacteriol.	180 (6), 1512-1524 (1998)
ACCESSION	U05170	AY056039	L03407	L13114	M33798	U04359
VERSION	L05170.5	GI.32306886				
KEYWORDS						
ORGANISM	Acinetobacter sp. ADP1					
REFERENCE	1	(bases 1 to 45324)				
AUTHORS	Hartnett,C., Neidle,E.L., Ngai,K.L. and Ornstom,L.N.					
TITLE	DNA sequences of genes encoding <i>Acinetobacter calcoaceticus</i> protocatechuate 3,4-dioxygenase: evidence indicating shuffling of genes and of DNA sequences within genes during their evolutionary divergence					
JOURNAL	J. Bacteriol.	172 (2),	956-966	(1990)		
MEDLINE	90130333					
PUBMED	2298704					
REFERENCE	2	(bases 1 to 45324)				
AUTHORS	Dimarco,A.A., Averhoff,B.A., Kim,E.E. and Ornstom,L.N.					
TITLE	Evolutionary divergence of poba, the structural gene encoding p-hydroxybenzoate hydroxylase in an <i>Acinetobacter calcoaceticus</i> strain well-suited for genetic analysis					
JOURNAL	Gene	125 (1),	25-33	(1993)		
MEDLINE	93194074					
PUBMED	8449410					
REFERENCE	3	(bases 1 to 45324)				
AUTHORS	Dimarco,A.A., Averhoff,B. and Ornstom,L.N.					
TITLE	Identification of the transcriptional activator poba and characterization of its role in the expression of poba, the structural gene for p-hydroxybenzoate hydroxylase in <i>Acinetobacter calcoaceticus</i>					
JOURNAL	J. Bacteriol.	175 (14),	4499-4506	(1993)		
MEDLINE	93322329					
PUBMED	8331077					
REFERENCE	4	(bases 1 to 45324)				
AUTHORS	Hartnett,G.B. and Ornstom,L.N.					
TITLE	Acquisition of apparent DNA alipage structures during extensive evolutionary divergence of pbad and pabd genes encoding identical catalytic activities in <i>Acinetobacter calcoaceticus</i>					
JOURNAL	Gene	142 (1),	23-29	(1994)		
MEDLINE	94237485					
PUBMED	8181753					
REFERENCE	5	(bases 1 to 45324)				
AUTHORS	Kowalchuk,G.A., Hartnett,G.B., Benson,A., Houghton,J.E., Ngai,K.L. and Ornstom,L.N.					
TITLE	Contrasting patterns of evolutionary divergence within the <i>Acinetobacter calcoaceticus</i> pba operon					
JOURNAL	Gene	146 (1),	23-30	(1994)		

QY	1483	TTGTGGGGCTGAGAAAGAGCG---TCGACGCCGAGAGTGCTTCCAGTGAGCCGGTT	1539	MEDLINE	94341565	
Db	1603	TATGAGGCTTTTAAAGCTGGCGCAATATTTCTTGACAGAGATATCTCAACACCTTAAAGTTC	1662	REFERENCE	8063101	
QY	1540	CGGGCGCTGTTTCTGACTGGCTCAACGACATCGATCGAAGCAACTGGCAATGCCAGT	1599	AUTHORS	6	(bases 1 to 45324)
Db	1663	CACACATGTTTCCGCCAATTTTAAACGACTTTTAAAGATGCGACTGGCAGCTCAAT	1722	TITLE	Elemore,D.A. and Ornstom,L.N.	The pca-pob supraoperonic cluster of <i>Acinetobacter calcoaceticus</i> dehydrogenase
QY	1600	CGCATCATGTGGTGAAGGCTCCTCGATACGCCGCCGTGATGATGAAGGGCGAGTCACT	1659	JOURNAL	J. Bacteriol.	176 (24), 7659-7666 (1994)
Db	1723	ACAGTCTCAATCTTTATTTAATGACGAGCCACCTCAGTTAGTCCGGCGAAGTGACC	1782	MEDLINE	95095936	
QY	1660	GACAAAGGCTCGATACACGAGCGCTGTTTTCATAGCGCGTCCGCCAAGTTATGCG	1719	REFERENCE	7	(bases 23320 to 27131)
Db	1783	GATAAAGCAACCTCAATCAAGCAGATTACCAAGCGTGGCGCTTTATTTATGATGAG	1842	AUTHORS	Elemore,D.A. and Ornstom,L.N.	Unusual ancestry of dehydratases associated with glutamate catabolism in <i>Acinetobacter calcoaceticus</i>
QY	1720	CTGTATC 1726		JOURNAL	J. Bacteriol.	177 (20), 5971-5978 (1995)
Db	1843	CTTTATC 1849		MEDLINE	96011389	
RESULT 12				AUTHORS	7592351	
ACCPAOP/c				REFERENCE	8	(bases 1 to 45324)
LOCUS	ACCRAP	45324 bp	DNA	linear	Gerischer,U., Segura,A. and Ornstom,L.N.	pcaU, a transcriptional activator of genes for protocatechuate utilization in <i>Acinetobacter</i>
DEFINITION	Acinetobacter sp. ADP1	dca-pca-qui-pob-hca	supraoperonic gene		J. Bacteriol.	180 (6), 1512-1524 (1998)
ACCESSION	U05170	AY056039	L03407	L13114	M33798	U04359
VERSION	L05170.5	GI.32306886				
KEYWORDS						
ORGANISM	Acinetobacter sp. ADP1					
REFERENCE	1	(bases 1 to 45324)				
AUTHORS	Hartnett,C., Neidle,E.L., Ngai,K.L. and Ornstom,L.N.					
TITLE	DNA sequences of genes encoding <i>Acinetobacter calcoaceticus</i> protocatechuate 3,4-dioxygenase: evidence indicating shuffling of genes and of DNA sequences within genes during their evolutionary divergence					
JOURNAL	J. Bacteriol.	172 (2),	956-966	(1990)		
MEDLINE	90130333					
PUBMED	2298704					
REFERENCE	2	(bases 1 to 45324)				
AUTHORS	Dimarco,A.A., Averhoff,B.A., Kim,E.E. and Ornstom,L.N.					
TITLE	Evolutionary divergence of poba, the structural gene encoding p-hydroxybenzoate hydroxylase in an <i>Acinetobacter calcoaceticus</i> strain well-suited for genetic analysis					
JOURNAL	Gene	125 (1),	25-33	(1993)		
MEDLINE	93194074					
PUBMED	8449410					
REFERENCE	3	(bases 1 to 45324)				
AUTHORS	Dimarco,A.A., Averhoff,B. and Ornstom,L.N.					
TITLE	Identification of the transcriptional activator poba and characterization of its role in the expression of poba, the structural gene for p-hydroxybenzoate hydroxylase in <i>Acinetobacter calcoaceticus</i>					
JOURNAL	J. Bacteriol.	175 (14),	4499-4506	(1993)		
MEDLINE	93322329					
PUBMED	8331077					
REFERENCE	4	(bases 1 to 45324)				
AUTHORS	Hartnett,G.B. and Ornstom,L.N.					
TITLE	Acquisition of apparent DNA alipage structures during extensive evolutionary divergence of pbad and pabd genes encoding identical catalytic activities in <i>Acinetobacter calcoaceticus</i>					
JOURNAL	Gene	142 (1),	23-29	(1994)		
MEDLINE	94237485					
PUBMED	8181753					
REFERENCE	5	(bases 1 to 45324)				
AUTHORS	Kowalchuk,G.A., Hartnett,G.B., Benson,A., Houghton,J.E., Ngai,K.L. and Ornstom,L.N.					
TITLE	Contrasting patterns of evolutionary divergence within the <i>Acinetobacter calcoaceticus</i> pba operon					
JOURNAL	Gene	146 (1),	23-30	(1994)		

QY	1483	TTGTGGGGCTGAGAAAGAGCG---TCGACGCCGAGAGTGCTTCCAGTGAGCCGGTT	1539	MEDLINE	94341565	
Db	1603	TATGAGGCTTTTAAAGCTGGCGCAATATTTCTTGACAGAGATATCTCAACACCTTAAAGTTC	1662	REFERENCE	8063101	
QY	1540	CGGGCGCTGTTTCTGACTGGCTCAACGACATCGATCGAAGCAACTGGCAATGCCAGT	1599	AUTHORS	6	(bases 1 to 45324)
Db	1663	CACACATGTTTCCGCCAATTTTAAACGACTTTTAAAGATGCGACTGGCAGCTCAAT	1722	TITLE	Elemore,D.A. and Ornstom,L.N.	The pca-pob supraoperonic cluster of <i>Acinetobacter calcoaceticus</i> dehydrogenase
QY	1600	CGCATCATGTGGTGAAGGCTCCTCGATACGCCGCCGTGATGATGAAGGGCGAGTCACT	1659	JOURNAL	J. Bacteriol.	176 (24), 7659-7666 (1994)
Db	1723	ACAGTCTCAATCTTTATTTAATGACGAGCCACCTCAGTTAGTCCGGCGAAGTGACC	1782	MEDLINE	95095936	
QY	1660	GACAAAGGCTCGATACACGAGCGCTGTTTTCATAGCGCGTCCGCCAAGTTATGCG	1719	REFERENCE	7	(bases 23320 to 27131)
Db	1783	GATAAAGCAACCTCAATCAAGCAGATTACCAAGCGTGGCGCTTTATTTATGATGAG	1842	AUTHORS	Elemore,D.A. and Ornstom	

REFERENCE 17 (bases 1 to 45324)
AUTHORS Smith,M.A., Young,D.M. and Ormston,L.N.
TITLE Direct Submission
JOURNAL Submitted (12-SEP-2001) MCDB, Yale University, P.O. Box 208103, New Haven, CT 06520-8103, USA
18 (bases 1 to 45324)
AUTHORS Smith,M.A., Young,D.M. and Ormston,L.N.
TITLE Direct Submission
JOURNAL Submitted (07-MAY-2002) MCDB, Yale University, P.O. Box 208103, New Haven, CT 06520-8103, USA
REMARK Sequence update by submitter
REFERENCE 19 (bases 1 to 45324)
AUTHORS Parke,D. and Ormston,L.N.
TITLE Direct Submission
JOURNAL Submitted (27-JUN-2003) MCDB, Yale University, P.O. Box 208103, New Haven, CT 06520-8103, USA
REMARK Sequence update by submitter
COMMENT On Jun 27, 2003 this sequence version replaced gi:20492839.
FEATURES
source
1. 45324
/organism="Acinetobacter sp. ADP1"
/mol_type="genomic DNA"
/strain="ADP1"
/db_xref="taxon:62977"
/note="Formerly Acinetobacter calcoaceticus"
/gene="dcaF"
complement (<1. .689)
/gene="dcaF"
/function="putative thiolase in catabolism of dicarboxylic acids"
/codon_start=1
/transl_table=1
/product="DcaF"
/protein_id="FAL09090.1"
/db_xref="GI:15812038"
/translation="MLNAVLYDGLRTPFGHAGELASIRPDDIAGLVIOQLIEKTGVA
GADIIDVYFDGTDQAGEDSRNIRHALLAGLPVPGQVNRILCSGLAAIIDSARA
ITGEGDLVJAGCSWSRAPFVWGKASYSRDAKIYDTTIGTFPNKKIIVQYGGH
SMPEGTGNVAVVYGISREQADLPAAQQAQKQALBEGFPAGEITAVEVGKKLPK
QVTEDEHPRPS"
complement (710. .2227)
/gene="dcaH"
complement (710. .2227)
/gene="dcaH"
/function="hydroxyacyl-CoA dehydrogenase in catabolism of dicarboxylic acids"
/codon_start=1
/evidence=experimental
/transl_table=1
/product="DcaH"
/protein_id="FAL09091.1"
/db_xref="GI:15812039"
/translation="MTHPIKKAIIIGVWVSGIAQIQAQSGHITVYIDAKAQAQQA
KQDLATPQKLDLKNKITTAYDAANAHLIANELHDLKCDLIVEIVRLDKSL
MSQLEAIPEPTTILASNTSLSTIALASNKHBERVAFNPVPMKVVEVIGKSL
TDPKHIETLNQLSRVLGHPRVAKDTPGFIINHAGRAYGEALAKIENETDISEDR
IILDGVRFMQPELMDLTGLDVSHPMESIYHOYVEAARYRNSLTOKMLAKQSR
KVGQGFVYRTSGTSETSAKVAERLLTVPRKWIADFEEDKOLLINLYTHNIOLDY
GAKPQADSLCLACGSEDTTHALRLNVPNPAHVAIDMLYGIETKRIMSLITETVY
SHAHSITFINDGAMVSTIGSIGFVAQRILIANVINAGCDIAQQAISVDINAVRIG
LGVPFPIEWGDISNKIILLNRIALTSDPRYRPSPLQRRVALNPLPTT"
complement (2276. .3037)
/gene="dcaC"
complement (2276. .3037)
/gene="dcaC"
/note="putative ketoacyl-CoA reductase"
/codon_start=1
/transl_table=1
/product="DcaC"
/protein_id="FAL09092.1"
/db_xref="GI:15812040"
/translation="MARKVLIITAGSGIGICIAEAPLIKQGDYFVCDINDSHQPRQ

Query Match 25.0%; Score 442.8; DB 1; Length 45324;
Best Local Similarity 55.2%; Pred. No. 2e-95;
Matches 930; Conservative 0; Mismatches 747; Indels 9; Gaps 3;
DYPQLHTFCCDLQGPYRYVQAMPDTAVAOQLGDLILNNTGSGPTVAEELSPFEWTO
VLIININSGFTLVTKLAIPLYKQSDAGVILNISIAKMGYFPLAISTSKMVGFTK
TISMEIGAFIRVNALIPGAVDERVORVQAADSMNISIEDVTENALANQSIKHFV
NPKHIADLICFIPLASDSGRISQIILPIDDKQLT"
46 GAGCGTCGAGCATTTGGCGTAAAGACCCGTCAGAACCAACCTGCGTGCAGAGCGC 105
Db 39842 GATCGTTGATTCATTTTGTCTAAAGTTAGACCCCTACATACCTTTGCGCTAAACGCAAT 39783
Qy 106 GCAATGGGGAATGAGCGTCGTATCAGCTACGCGGAAATGTTCCACAAGTCGCGGCATC 165
Db 39782 ACACAGACGAATGATGATCAGGCTCAGTTATGAGAAACCTTGCAAGCGCTTGGAATAT 39723
Qy 166 GCAAGAGCTTGCTTCTTACGGAATACGAGAGAGCGCTCGCTTATCGTCTGGA 225
Db 39722 GCACAAAGCTTTCATCAGCGCACTTAAGCCAAAGAACGACCTTGGTAATGTATCGGGA 39663
Qy 226 AATGACCTGGAATCTTACGCTGCGCATTTGGGGCTATGTATGCGGGCATTCCTATTCG 285
Db 39662 AATGATCTTGAACTTAACGCTTTCGATGAGGCGCGCATGTTGGCAGGTATACATTTTCT 39603
Qy 286 CCGGTGCTCTGCTTATTCATCTGCTGCGCAAGATTTGGCGAAGCTGCTCATCTGA 345
Db 39602 GCTATTTTCGACAGCTATCAGTATATCAGATTTTGGCAATTAATAAGTGTGTT 39543
Qy 346 GGTCTTCTGCAACCGGAGCTGCTTTCGTCGCGGATGACACCTTTCCAGCGCGCAAT 405
Db 39542 GATGTGCTACCCCAAGCGATGGTGTATGCAATGATGACAACTTTTGTCTAAAGCAT 39483
Qy 406 GAGACCATTCGCGCGGACGACGCGCATTTCTACCTGAGGCGCAATTTGGCGGCGG 465
Db 39482 CAAGCTGTACGCAAGTCTATTTGAATTTGTACCAATTAAGCATTTATAGTGATCA 39423
Qy 466 CGCAGCGTAGTTTGAACAGCTGCTGAGACAGCTGTGGGATGAGCGAGATTAATGCC 525
Db 39422 GTCTGTACATCGTTTCACTTATTAATGATACCCCTGTGACCGATGTTCAAGAACAT 39363
Qy 526 TTTGGCGCACTGCGCCCATATGATTTGCCAATTTCTTTTCACTTGTGCTCTACCAA 585
Db 39362 GAAACCTGATATGAC--ATCAGATTTGCCAATTTTATTCATTCGCGTTGACCAAG 39306
Qy 586 CTGCTTAAGCGGTGCGCACTACTCAAGCAATGCTGCGCCATGACAGATGCTCTG 645
Db 39305 CTTCGAAGGCTTTTCGACCAACATTTGATGCTGTATCAATCAACAAATGCTACT 39246
Qy 646 CAACCTTTCCGGTTTGTGTGAAGAGCGCGGTGCTGTGAGCTGTTCCGTGGAAC 705
Db 39245 CAACATTTTCAGAAATTTGAAGACACACCGCGCTGTCTTACTGAGCTGTGTCTTGGCAT 39186
Qy 706 CAACCTTGGCGGCGGACCCCAACATTCGCGATGCTGTGTGTAACAAGCGCGACGTA 765
Db 39185 CATACATTCGTGTGAAGCCCAATGTTGGATGCGCTTATTAACGCGGAAACATTTAC 39126
Qy 766 CTGAGAGCGGTAAACCAACCGCCCAAGGTTGCGCGAAGCGCTTGCAACTGAGCGAA 825
Db 39125 ATGATGACGATAGCCCGTTCGCGGAAATTTGATGAAACGATTTGAAATTAAGA 93066
Qy 826 ATCTCTCCACTGCTACTGCTGTCGCGGAAAGGCTGGAGAGAAATTAAGTGGTGCCTT 885
Db 39065 ATTTCAACGACGCTTATCTTAATGTACCAAGGCTGGGAGAAACTCACCGATGACCT 39006
Qy 886 GAGCGAGACATACCTCGCGGAAAGCTTCTTGTGCTGATGAAGCTGTTCTTTCGCG 945
Db 39005 GAAAGAGATGCGCAATCCAGAGATCGCTTCTTGCAACGTTAAATTTGTGTTTGTCT 38946
Qy 946 GCGGCTGGGTTGCGCAAGGAGATCTGGAGATGTTTGAACCGGCTGCTGAACAGACTGT 1005
Db 38945 GGTGCTGCAATTTCTGAAGAGGCTGGAACCGACTGATTAATTTGACAGGACATTTGC 38886

Qy	1006	GGTAGCGCATTCGATGATGCGGCTCTGGGATGACGAGATCGTCTCTGCACT	1065
Db	38885	GGTGAATAAATTCGATCATGATGCTTGGGATGACGAAAGGACCATCTGTGCGC	38826
Qy	1066	TTTACACCGGACCGCTGCTGATGCTGTTCATTTGGGCTGCGAGCGCTGCGAG	1125
Db	38825	TTTACACGAGCCGACGCGATGATGCTGTTCATTTGGGCTGCGAGCGCTGCGAG	38766
Qy	1126	GTCAAGCTGCTTCCGCTGATGAGAAATTGGAAGGCGTTTCATGCTCCGACATG	1185
Db	38765	ATCAAAATAGTTCCTTATGATGACAACTGAAATTTGTGTTCTGCGACAGCATG	38706
Qy	1186	AGGCGCTACTGGG---TGCCTGAAACAAATGCCCAAGCGTTCGACGAGAGGCTAT	1242
Db	38705	AAAGTTTACTGGCCCTTAAGGCCGACGACAAACACATTTTGAATGAAGCGCTTT	38646
Qy	1243	TACTGCTCCGCTGATGCTCATMAATTGGCAGATCTGCGCATCTCAGAAAGTCTGATG	1302
Db	38645	TTCATATCTGGCGATGCGGATGCTGTGCTGATCAAAACGACCTGACAGGGCTGATG	38586
Qy	1303	TTTGAACGCTGAAATTCCTGTAAGACTTCAGCTGCTCAGGGGATTTGTCAAGCTTGGG	1362
Db	38585	TACGACGCGCGTATTTGCAAGAAATTTCAACTCAACGCGGTACCTTTGTAATGTGGC	38526
Qy	1363	CCATTCGCGACGCGCGGCTCTGGAAGCGGCTCTTACGCTCTGACGTAGTGTGCT	1422
Db	38525	ACTTTACGCAACAGGACATCATTCATTAAGTTATTTGTTGATTCAGAAATTTTGAATTAAC	38466
Qy	1423	GCTCTGATCGTGAATCCTTGAATGCTGCTGTTTCGCGCTCTTCTGACATGCGGCTCC	1482
Db	38465	GGTTCAATCTAAATGCGATTTGCTTCTGTTTCTTGAAGTTGATGCTGTGCAAGAT	38406
Qy	1483	TTGTGCGGGCTAGAAAAGAGGCGTCCGACGCGAGGTGCTGCGCAATGAGCCG---GTT	1539
Db	38405	TTTGTGCTGCTGAATATGAATAAATCATAGCGCAATATATCTTGAACACCCCTTAAAGTA	38346
Qy	1540	CGGCGCTGTTTGTGCTGCTGCTCAACGACTCAATCGAAGAGCACTGGCAATGCCAGT	1599
Db	38345	CAGCAGTGTGTTTGCACAGTTTATTCATGACAAACAAATGTCACGCGGACGTTCAAT	38286
Qy	1600	CGCATCATGTGGTAGGCTCTCTGATACGCGCGCTGATGTTAAAGGCGAGGTCACT	1659
Db	38285	CGGGTATCAATGTTGATTTAATGACTGAGACACACAGCTGATGAGGTAAC	38226
Qy	1660	GACAGGCGCTGATCAACGAGCGCTGTTTTCATGAAAGCGGCTGCGGAAAGTTGATGCG	1719
Db	38225	GATTAAGGCAATTAATTAATGAGTACATCTTAACGTCGTGCTGACGTGTGAAGAA	38166
Qy	1720	CTGTAT 1725	
Db	38165	CTGTAT 38160	

RESULT 13

CR543861.17/c

Sequence split into 36 fragments LOCUS CR543861 Accession CR543861

Fragment Name	Begin	End
CR543861_00	1	110000
CR543861_01	100001	210000
CR543861_02	200001	310000
CR543861_03	300001	410000
CR543861_04	400001	510000
CR543861_05	500001	610000
CR543861_06	600001	710000
CR543861_07	700001	810000
CR543861_08	800001	910000
CR543861_09	900001	1010000
CR543861_10	1000001	1110000
CR543861_11	1100001	1210000
CR543861_12	1200001	1310000
CR543861_13	1300001	1410000
CR543861_14	1400001	1510000

Qy	CR543861_15	1500001	1610000
Db	CR543861_16	1600001	1710000
Qy	CR543861_17	1700001	1810000
Db	CR543861_18	1800001	1910000
Qy	CR543861_19	1900001	2010000
Db	CR543861_20	2000001	2100000
Qy	CR543861_21	2100001	2210000
Db	CR543861_22	2200001	2310000
Qy	CR543861_23	2300001	2410000
Db	CR543861_24	2400001	2510000
Qy	CR543861_25	2500001	2610000
Db	CR543861_26	2600001	2710000
Qy	CR543861_27	2700001	2810000
Db	CR543861_28	2800001	2910000
Qy	CR543861_29	2900001	3010000
Db	CR543861_30	3000001	3110000
Qy	CR543861_31	3100001	3210000
Db	CR543861_32	3200001	3310000
Qy	CR543861_33	3300001	3410000
Db	CR543861_34	3400001	3510000
Qy	CR543861_35	3500001	3598621

Continuation (18 of 36) of CR543861 from base 1700001 (CR543861 Acinetobacter sp. ADP1 c

Query Match 24.9%; Score 441.2; DB 1; Length 110000;
 Best Local Similarity 55.1%; Pred. No. 4.9e-95;
 Matches 929; Conservative 0; Mismatches 748; Indels 9; Gaps 3;

Qy	46	GAGGCTTCGACGATTTGGGCTTAAGACCCGTCAGAACAAACCTGCTGCTCCAGGGCG	105
Db	33693	GATGCTTGATTCATTTTGCTTAAAGTTTAAAGCTTACCATCTTTGCGCTTAAACGCAAT	33634
Qy	106	GCAATGGGGAATGCGGCTGATCAGCTACGCGAAATGTTCCACAAAGTCCGCGCATC	165
Db	33633	ACACAGACAAATGATGATCAGGCTCAGTTATGCAAACTTGCACACGCTTGCAATTT	33574
Qy	166	GCAAGAGCTTGTCTCTTACGACATACGCAAGAGCGCTCGCTTATGCTCTGGA	225
Db	33573	GCACAGCTTTGATCAGGCCCACTTAAGCCAAAGACCTTTGTAATTTTATCGGA	33514
Qy	226	AATGACCTGAAATCTTCACTGCTGCAATTTGGGCTATGTAATGCGGCAATTCCTATTGC	285
Db	33513	AATGATCTGAACACTTAAACCTTACGCTTTCATGCGGCGATGTTGGCAGGTGTAACATTTCT	33454
Qy	286	CCGGTCTCTGCTTATTAATCACTGCTGCGCAAGATTTGGCGAAGCTGCGTCATGTA	345
Db	33453	GCTATTTGCGCAGCTACTACGTGATCTCAAGATTTGGCAATTAACATGCTTT	33394
Qy	346	GGTCTTTCGCAACCGGAGCTGCTTTGTGCTGCGATGACACACTTTCCAGCGGCAATT	405
Db	33393	GATGTGCTACCCGAGCATGCTGTAATGCAATGATGACAAAGCTTTTGCATAAGCAATT	33334
Qy	406	GAGACATTTTCGCGGAGCAGCTGCGCCGCAATCTTCACTGAGCGCAATTTGCGCGGCGG	465
Db	33333	CAAAAGCTGTACGAAAGTATCATATTAATGTAACCAATTAAGGCATTAATAGTGAACAA	33274
Qy	466	CGACGCTAGTTTGAACACCTGCTGAGACGCTGCTGAGATTTGAGGAGATTAATGCC	525
Db	33273	GTCGTACATCGTTTCACTGCTTTATTAATGATACCCCTGTGACCGATGTTCAAGAACATTAAT	33214
Qy	526	TTTGGGCAACTGGCCCGATAGCATTTCCAAAGTTCTTGTTCATCTTGTGCTTACCAA	585
Db	33213	GAAAGCTAATAGC---ATCAGATTTGCAAGTTTATTTTATTCATCGGTTGACCAAG	33157
Qy	586	CTGCGTAAGCGGTGCGGACTACTCAGCGAATGCTTGCAGCAATCAGCAGATGCTTCTG	645
Db	33156	CTTCCAAAGGCTGTTCGACACACATTTATGCTGTGATCATCAATCAACAAATGCTACTT	33097
Qy	646	CAAACTTCCCGTTTGTGTAAGAGCGCGGCTGCTGATGAGCTGTTGCGCTGAGAC	705
Db	33096	CAAACTTTCAGAAATTTGAAGACACCGCGCTGCTTACTGACGTGTTGTCTTGGCAAT	33037
Qy	706	CACACCTTCCGCGGACGCAACATCGCATGCTGTTGTGAACGCGGACGCTACTAC	765

Db 33036 CATACATTCGGTGAAGCCACAAATGTTGGATTGCGCTTTATTAACGCGGACCATTTAC 32977
QY 766 CTGAGACGGGTAAACCAACCCGCCAAGGGTTGCGCGAGCGCTTGCAACTTGAGGAA 825
Db 32976 ATGATGACGGTAAAGCCGTTCCGGGAAAAATTGATGAAACGATTGAAATTTAAAAAGAA 32917
QY 826 ATCTCTCCCACTGCGTACCTCACTGTGCGGAAAAGGCTGGAGGAATTAAGTGGTGCCTT 885
Db 32916 ATTTACACGACGGTTTATCTTAATGTAACCAAGGCTGGGAGAACTCACCGATGACCTT 32857
QY 886 GACGAGACATACCTCGCGCCAAAGCTTTCTTGCTGCGATGAAGCTGTCTTCTTCGCG 945
Db 32856 GAAAAAAGATGCCAATCAAGATCCCTCTTGCAACGCTTAATAATTTTGTGTTTGTCT 32797
QY 946 GCGGCTGGGTGTCGCAAGGAGTCTGAGATCGTTTGAGACGGGCTGCGTGAACAGCACTGT 1005
Db 32796 GGTGCTGCAATATCTGAAGAGGCTGGAAACGACTGATTAATAATTGACAGGACATTTGC 32737
QY 1006 GGTGACCGCATTCGATGATGCGGGTCTGGGCAATGACGAGACTGCTCTTCTGCACT 1065
Db 32736 GGTGAAAAAATCCGATCATGATGCGCTTGCGCATGACGAAACGCGACCATGTCGCGC 32677
QY 1066 TTATACACCGGACCGCTGTCGATGCGCTGTGTTACATTGGGCTGCGACGCGCTGCGAG 1125
Db 32676 TTATACACAGGCTCCAGGCTAATGCGCTGTTTATTTGGCTAACCTGCTCGGGCTGTGA 32617
QY 1126 GTCAAGCTCGTTCGGGTGCAATGGAAGGAGGCTTTCATGAGTCCGACGTCATG 1185
Db 32616 ATCAATTAATGATCCCTATGCTGACAACTTGAATTTGTGTGTGTCGCAAGCAGCTCATG 32557
QY 1186 AGCGGCTACTGCGG---TGCCTCTGAACAAATATGCCAAGCGTTCGACGAGAGGCTAT 1242
Db 32556 AAGGTTACTGCGCGCTTAAGCGCCGACACGACAAACAAATTTTGTATGAAAGCGCTTT 32497
QY 1243 TACTGCTCCGCTGATGCAATCAATTTGGAGATCTGCGCATCTCAAGAAAGTCTGATG 1302
Db 32496 TTTCATATGCGGCATCGGATCGCTGCTGTCGATCAACGACCTCGACAGGGCTGATG 32437
QY 1303 TTATGACGCTGCAATTCCTGAAGACTTCAAGCTGCTCCAGGGGTATTTTGCAGCGTTGG 1362
Db 32436 TACGAGCGCGCTATTTGCAAGAAATTTCAACTCAACGCGGTACCTTTGGAATGTGGC 32377
QY 1363 CCATTGCGCACGCGCGCGGTTTGAAGCGCGCTTTACGTCCTGACGATGCTGTTCT 1422
Db 32376 ACTTTACGCAACAAAGCACTCATTTCAAGGTAATTTGTGATTCAGATGTTGTATTAAC 32317
QY 1423 GCTCCGATGCTGAATGCTTGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1482
Db 32316 GGTTCAAATCTTAATGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 32257
QY 1483 TTGTCCGCGCTAGAAAAAGAGCGTGGAGCGCGGAGTGTCTTCCAGTGAAGCGG---GTT 1539
Db 32256 TTGTGCTGCTGATTAATGAAAAATCATAGGCAATAGATATCTTGCAACACCTTAAGTA 32197
QY 1540 CCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1599
Db 32196 CAGCACTGCTTTCGACAGTTTAAATTCAGTACAAAGATGCTCCAGGCACTTCAAAAT 32137
QY 1600 CGCATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1659
Db 32136 CCGGATTCATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 32077
QY 1660 GACAAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1719
Db 32076 GATTAAGGCAATTTAATGAGATTAATCATCTTAAAAAGTGTGCTGCTGCTGCTGCTG 32017
QY 1720 CTGTAT 1725
Db 32016 CTGTAT 32011

RESULT 14
AX024534

LOCUS AX024534 1203 bp DNA linear PAT 15-SEP-2000
DEFINITION Sequence 45 from Patent DE19850242.
ACCESSION AX024534
VERSION AX024534.1 GI:10184691
KEYWORDS
SOURCE
ORGANISM
Pseudomonas sp. HR199
Pseudomonas sp. HR199
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE
1 Overhage, J., Priefert, H., Rabenhorst, J. and Steinhuechel, A.
Patent: DE 19850242-A 45 04-MAY-2000;
HAARMANN & REIMER GMBH (DE)
FEATURES
source
1..1203
Location/Qualifiers
/organism="Pseudomonas sp. HR199"
/mol_type="unassigned DNA"
/db_xref="taxon:86003"
342..>740
/note="unamed protein product; N-Terminus (AMINO ACID 1
bis 133) codierender Teil des
Ferulasaure-CoA-Synthetase-Gens (fcs) "
/codon_start=1
/protein_cable=11
/protein_id="CAC08976.1"
/db_xref="GI:10184692"
/translation="MRSLEALPLPFGILRLERLHNAKTRPEQTCVAARANGSMRIS
VAEMFNVAIALISLLPYGLSABRPILITVSGDLEHLQAFGAMVAGIRYCPVSPAYS
LSQDLAKIRHIVGLQPLGVPAADAAPQR"
<741..827
/note="unamed protein product; C-Terminus (AMINO ACID 562
bis 589) codierender Teil des
Ferulasaure-CoA-Synthetase-Gens (fcs) "
/codon_start=1
/transl_table=11
/protein_id="CAC08977.1"
/db_xref="GI:10184693"
/translation="AVLQWRSAKVDALYRGEDQSMLEDEATL"

ORIGIN

Query Match 22.9%; Score 404.6; DB 6; Length 1203;
Best Local Similarity 91.6%; Pred. No. 3e-86;
Matches 428; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
QY 1 ATGCGTTCTCTCGAGCGCTTCCCTTCCGCGGTGCAATCTTGAGCGTCTCGAGCAT 60
Db 342 ATGCGTTCTCTCGAGCGCTTCCCTTCCGCGGTGCAATCTTGAGCGTCTCGAGCAT 401
QY 61 TGGGCTAAGACCCGCTCAAGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120
Db 402 TGGGCTAAGACCCGCTCAAGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 461
QY 121 CGTCTGATAGCTAGCGGGAATGTTCCAAACGTCGCGGCCCATGCGACAGACTTGCTT 180
Db 462 CGTCTGATAGCTAGCGGGAATGTTCCAAACGTCGCGGCCCATGCGACAGACTTGCTT 521
QY 181 CCTTAAGCACTATCGGACAGGCTCCGCTGCTTATGCTGCTGGAATGACCTGGAACAT 240
Db 522 CTTTACGACTATCGGCAAGCGTCCGCTGCTTATGCTGCTGGAATGACCTGGAACAT 581
QY 241 CTTTACGACTATCGGCAATGCTATGATGCGGCGATTCCTTATGCGCGGTCTCTGCT 300
Db 582 CTTTACGACTATCGGCAAGCGTCCGCTGCTTATGCTGCTGGAATGACCTGGAACAT 641
QY 301 TATTGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360
Db 642 TATTGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 701
QY 361 GGAAGTCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
Db 702 GGAAGTCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 761
QY 421 GACGACGTGCCGCAATTTTCACTCGAGGCGAATTTGGCGGCGCGG 467

Db 762 GCCAAGTGTATCGGTGATGATCATCATCATCGCG 808

RESULT 15
AE008067/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
ORGANISM

10173 bp DNA linear BCT 28-MAY-2004
Agrobacterium tumefaciens str. C58
of 254 of the complete sequence.
AE008067 AE007269
AE008067.1 GI:15156484

Agrobacterium tumefaciens str. C58
Agrobacterium tumefaciens str. C58
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium;
1 (bases 1 to 10173)
Hinkle,G., Slater,S.C. and Goodner,B.
Complete Genome Sequence of Agrobacterium tumefaciens C58
(Rhizobium radiobacter C58), the Causative Agent of Crown Gall
disease in Plants
Unpublished
2 (bases 1 to 10173)
Hinkle,G., Slater,S.C. and Goodner,B.
Direct Submission
Submitted (14-AUG-2001) Cereon Genomics, 45 Sidney Street,
Cambridge, MA 02139, USA

FEATURES
source

1..10173
/organism="Agrobacterium tumefaciens str. C58"
/mol_type="genomic DNA"
/strain="C58"
/isolate="Cereon"
/db_xref="taxon:176299"
/complement (100..1263)
/gene="AGR_C_2609"
/complement (100..1263)
/gene="AGR_C_2609"
/note="branched-chain amino acid ABC transporter,
periplasmic amino acid-binding protein - Deinococcus
radiodurans (strain R1)"
/codon_start=1
/transl_table=11
/product="AGR_C_2609p"
/protein_id="AAK87205.1"
/db_xref="GI:15156485"
/translation="MGKTMKRIINTWMAVAALTTAGSAFADVTIGTTPPTGPA
ALGIPLKOSIEFPADIDGKLVILDDGDSVATTNRRASDDKVDVIGSALT
PAIAVAGVANETEPHFALSPITPAATKKNFMPOFSTVLKSMFARMQKXGRV
AFGPAFDSYGQGVDAKATGKXGVLVAEERYARADTSVGOVLKSLAARPAIPI
GASGTGAALPQIGVRRGRGKRYVVSAGVTPPLIAGAVDNTVPASGVPAVEDQ
PDSAPTKQESGLATLKAPEGKFGENTRQFSGHWVDWVILKAVPALKKAKEGTAEF
RAALRDAIETSGPYNATHGYEFTLANHYGLDGRGVLLTPADGKFFVL.PDNK"
/complement (1317..2960)
/gene="AGR_C_2611"
/complement (1317..2960)
/gene="AGR_C_2611"
/note="probable acyl-CoA dehydrogenase - Deinococcus
radiodurans (strain R1)"
/codon_start=1
/transl_table=11
/product="AGR_C_2611p"
/protein_id="AAK87206.1"
/db_xref="GI:15156486"
/translation="MKSQGMVASACSGDVSFSTQAAATMELLKATSGMRRIOGLR
GDCDDALVAAILNEARFATVLEPLATPDREGCTVVGVRVPTGPADEFRAHGD
GMIAVDVPERFGGALPLVQAACAPLFEFGCAALMAASGSTRACHLAETAEAGVA
DEWVPRLAGEWAATTCISRPDAGSDLARLSAARSGAMTRIGRKIMSRDHDMA
MRGHCLARTGETRTGSLPLVPEDEIDGKNGYVERIEKMGILHSPITALFD
GAAGLLGEGRGALQIPATIELMRLOTGQGGGLASALIDIAETVAGEQOQGDPAK
PAVALAHDPDRLOEARLETTLRAATLELATVNDLAMEPEKEBLRGLQLMSLWL
LPLVKNFGADAGFTVNOAIOIGAGYTEMPELEOYLARDARVMTIYEGTTGQALDL
LTRLWRGEGKGLDVLKRAKRESLADYRSDKTKAASVLDAPAESRRMAALENEPET

gene
CDS

GLWQANNVILNAWEAFAMMTVRVEGVADAQARRTGKAA"
/complement (2914..3687)
/gene="AGR_C_2613"
/complement (2514..3687)
/gene="AGR_C_2613"
/note="probable short-chain dehydrogenase PA2554
{imported} - Pseudomonas aeruginosa (strain PA01)"
/codon_start=1
/transl_table=11
/product="AGR_C_2613p"
/protein_id="AAK87207.1"
/db_xref="GI:15156487"
/translation="NNIBGAGALVTGAASGLGAIVARMILARGAAVTFIPRNGEAGK
LAAEIGKAVQGDVTDADAQAAIKVAASAKGLRILVNCAGIGTAGRILGREGPOL
GDEQVTRVNLICTFMNMLAAAHMAEREDBEDSDSDRONGVINTASVAFEGQICQ
AAYAAKSGIVSLALPAAEELARFRIRNVAVAGITPLPLQGLPDEVEESLAGQIPIH
PSRLGDPAEADDTVRPLIENDYNGEIVRILDAIRMQPR"
/complement (3684..5459)
/gene="AGR_C_2614"
/complement (3684..5459)
/gene="AGR_C_2614"
/note="(A0238746) feruloyl-CoA synthetase"
/codon_start=1
/transl_table=11
/product="AGR_C_2614p"
/protein_id="AAK87208.1"
/db_xref="GI:15156488"
/translation="MIRNDPLGPYPQKLNERRILHWCRAPERTMADROGPEPMRWY
SYAELDKIRRIQFLIDHLSVERPLVLSSENSIBHALVLAQHVGLASAAITPAY
ATSAIDLKADIIGQITPGWVFAEDATPFRALGEVDDGTPLVGLHNLGDSNTHF
FETLLETPBEAVDRAFDVAGPDVAKFLFETSGTGSPKAVIGTORHLGDSNIMLD
YGVREPEPVVDMAEPMNHTAAGNKVFNILVNGCTGYVIDRGRSPAOICGQTDNIRD
ISPFTVFNYPAGHEMLVOAMKRDALCRSFRRLKMLMAGAMAGHTDALTELISA
TVGAALVMAAGLSTETAPPSLCTEPQDPKGNIGIPAOGVTKLVPPDERVELRKG
PNITPGYWRNGELTAAAFDEEGFYRGDVPKFAVDPPRGFYFDGRMMEFNLQGT
WVAVGPRLAQLVMFAGLIRDAVITGENRAELGALVPPFIPALREHVRGSHLSDSEI
IRHPSVAQIVAKLSAHKOASGSAVRNRIILMEDALRPEKEGEVTDKGSINGRAVIL
HREKELVESLYGDPQVITVGRVAA"
/complement (5458..5610)
/gene="AGR_C_2616"
/complement (5458..5610)
/gene="AGR_C_2616"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_C_2616p"
/protein_id="AAK87209.1"
/db_xref="GI:15156489"
/translation="MISPPGRFSHATSGDSFTLIYVTSQSFTRTVRIIFCQKNDY
KRIISL".6461
/gene="AGR_C_2617"
/complement (5664..6461)
/gene="AGR_C_2617"
/note="enoyl-CoA hydratase (fad-1)"
/codon_start=1
/transl_table=11
/product="AGR_C_2617p"
/protein_id="AAK87210.1"
/db_xref="GI:15156490"
/translation="MSEKMTSEFVSVELRGDIALVGLANRPOKRNATSDRPFVEAIAAA
AAAEENARAAYIFRGHDFRCAGLDGEHHTKPIBEVRSRRKHAFVSOIERTHTWV
SAHAGVVGGSGLAEASTHRLVADDSAFPALPEGQKQVIGGGSVRIKRLGVAAMS
DMULTGRVASSAEABERNVLYOAVPREGALAAQELATTI AASGASVSNYAVINALPRI
QMAKEDGDFVSEFTSSFTNAPSEAEERLRAFLERKATRLKAPDITTA"
/complement (6511..7506)
/gene="AGR_C_2619"
/complement (6511..7506)
/gene="AGR_C_2619"
/note="(A011317) MetP protein"
/codon_start=1
/transl_table=11
/product="AGR_C_2619p"

```

/ protein_id="AAK87211.1"
/db_xref="GI:15156491"
/translation="MVFRLIEVAVSCIHNQVSLREKAMGMSFMDNSGADGVYRS
SNLWMSIEVMPRETKAIDIPNLIIDGTRVYIAHIDGPIADWSTARKLHDEGF
VNHLPARGISGNELEMDLKYQEGAGTYOALLGGGKOPAGSSGSDIMEITLF
DKMGFTHLHAGHPENRDIIDSSSRGDOLIMKODASRTDALATTQPAFAK
VVIEMAKRIGNMGVITLVHVGILGPTKLTLLKFAIACGVPGLKVLQKRALDLSLF
VVEPPTFAELIIEHKAANDSLIEHVIHIFLGGIKPSAEWMKQHASTGTLANVA"
gene
7739 .8473
/ gene="AGR_C_2620"
7739 .8473
/ gene="AGR_C_2620"
/ gene="AGR_C_2620"
/ note="probable transcription regulator PA4906 {imported}"
- pseudomonas aeruginosa (strain PAO1)"
/ codon_start=1
/ translation="M"
/ product="AGR_C_2620p"
/ protein_id="AAK87212.1"
/db_xref="GI:15156492"
/translation="MKQVDAATHGKRAVIELREKILSGELPGMRLEFVSTAEIDIS
KVPVEBALSLTEBGLNRLPGGFVVRFRFADVVDAIEVHGMEGTARLAERGV
SKVALEIDATVOQLDLCFGRVDVDFGVYALNRI FHHOLAALCGSEMRPREVRA
SSLVPSAPSAFLPDKANIGAPRRSLRGAQEHKAIYAAIVAREGAALAEVAHSHSTA
RNLEMYMREARBELIAQVPGALISD"
6623 .10044
/ gene="AGR_C_2622"
6623 .10044
/ gene="AGR_C_2622"
/ note="PROBABLE AMINOMETHYLTRANSFERASE (GLYCINE CLEAVAGE
SYSTEM T PROTEIN)"
/ codon_start=1
/ translation="M"
/ product="AGR_C_2622p"
/ protein_id="AAK87213.1"
/db_xref="GI:15156493"
/translation="MTMSSSLTEVLYKQYDNPVEMLRNKGIVYVYVAVGSEPMRS
EQARREBAVLFDQSHHMDLVEGPDARKPLATYHGINSPFNDLRAKHPVPTDG
HVIQDHIIFRERQDKFLLVGRAPTSWMLFCAAYGMWNRIRHDPSPRSEGRVW
THYRVOIQPDAKPIFEKNMGSPIDIKFHVDMINIGSKYQALHSGMSGALFVW
GYEDKSYLLSCILEAKQAGVNLVRCGRAYSTNLTJESGWIPLGPIYTGDMKLD
YDMLGADSYEATGAIQGSFVSENIEDYVNPBELGIDYIGMKKDPFGKSLKEMK
GMSHKKTKYFEMNAEDVNVITASAPRGEDHKWIDIPRANVASTGADILILGQKHYG
MSMPTQVSYNERCIIISLGIYVNDIOLGIVLTLVWGPDPGSAKSTERKQKEIRVRV
SLTPYAAERADNVADSWRTKKAS"
ORIGIN
Query Match 22.6%; Score 399.4; DB 1; Length 10173;
Best Local Similarity 52.8%; Pred. No.5.4e-85;
Matches 909; Conservative 0; Mismatches 806; Indels 6; Gaps 2;
QY 12 CGAGGCGCTTCTCCCTCCCGGCGGAATTCTTGAAGCGTCTCGACATTTGGGCTAAGAC 71
Db 5445 CGATCGGCTTTGGCCCTTATCCGCAAAATTGAATGAGCGTCTCTTCACTGGTGTGCTTC 5386
QY 72 CCGTCCAGAACAAACCTGCGTGTCTCCAGGGCGGCAAAATGGGAAATGGCGTGTATCAG 131
Db 5385 GGCAACCGGAGCGCAATGATGCGCCGACGTCAGGAGTCCGAGACCAATGGGCGCGGTCTAG 5326
QY 132 CTACGCGGAATGTTCCACACGTCGCGGCATCGACAGAGCTTCTTCTTACGGAAGT 191
Db 5325 CTACGCGGAGGCTCTGACAAAGTACGGGATACGGCCAGTCTTCTCTGATATACACT 5266
QY 192 ATCGGAGAGCGTCCGCTGCTTATCGTCTCTGGAATGACCTGGAAATCTTTCAGCTGCG 251
Db 5265 TTGCGTCGAGAGGCGCTGCTGCTTCTTCTGAAAATTCATGACATGACGCTGATAGGT 5206
QY 252 ATTTGGGGCTATGATGCGGCAATTCCTTATGCCCCGGTGTCTCTGCTTATCTAGTCT 311
Db 5205 GCTTGGGCGCCAGCATGTCGGTATCGCTCGCTGCAATCAACGCGGCTATGCTC---CAC 5149
QY 312 GTTCGCAAGATTTGGCGAGGTGCGACATCGTAGGCTTTCTGCAACCGGGAATGGGCTT 371
Db 5148 GTTCGCGCATCTCACAAACTCGCCGACATCGCGGCGAGATCAACCCGCGGATGGTTT 5089
```

```

QY 372 TGTGCGCATGAGCACTTTTCAAGCGCGCAATTGAGACCAATTCTGCGGAGCAGCGTGC 431
Db 5088 TGCCGAGATGAGGAGCGCTTTTCGCGGCGCGCTCGGCGAGGTTTTCAGACGAGTACACC 5029
QY 432 CGCAATCTTCATCTGAGGGAATTGGCCGGGGGGCC---ACGGTACGTTTTCAGACGCT 488
Db 5028 GCTGTGCGGCTCCCGCAACTCGCGGAATGCTTCCAAATACGTTTCAATTTTGAAACCTT 4969
QY 489 GCTGAGCAGCGCTGTGGGATTTGAGGCAATATGCTTTGGCGCAACTGGCCCCGATAC 548
Db 4968 GCTTGAAGAGAGCGGACCGAAGCGTTTATGCTGCTTTGATGCGGTGGCGCGGATAC 4909
QY 549 GATTCGCAAGTTCTTGTTCATCTTGGCTCTTACCAACTGCTTAAAGCGGTGCGCATAC 608
Db 4908 CGTGCGCAATTCCTTTTACGCTCAGGCGACGACTGGTGTGCGCAAGCGGTATCCAGAC 4849
QY 609 TCAGGCAATGCTGTGCGCAATCAGAGATGCTTCTGCAAACTTTCCCGTTTTCGTATA 668
Db 4848 GCMAAGCATGTTGTGCTCAATCAGGAATGATGCGCATTTGTATGCTACTTTTGCGCA 4789
QY 669 AGAGCGCGGCTGTGCTGGAAGTGTGCGCGTGAACACACACTTGGCGGAGCGACACA 728
Db 4788 GAGAGCGCGGCTGTGCTGATTTGGCGGCAATGAAACACACGCGCGGGAACAAGGT 4729
QY 729 CATCGGCAATGTTGTGTAACAAGCGCGGCAAGTACTTGAACGACGTTAAACCAACCC 788
Db 4728 GTTCAACCTGTCCTTATATAGCGGTACTTATATATGATTCGCGGCAAGCGCACGCC 4669
QY 789 CCAAGGCTTGGCGGAGCGCTTCGCACTTGAAGCGGAATCTCCACAGTGGTACTCAC 848
Db 4668 GGCGCAGATGCGCGACGACCTGACATCTCGGACATATGCGCCACTGTATTTTCAA 4609
QY 849 TGTGCGGAAGGCTGGAAGAAATGATGAGTGCCTCTTGAAGCAGACATGATCCCTGCGCA 908
Db 4608 CGTTCGCGTGCACAGAAATGCTGTCCAGGCAATGCGCAAGAGAGAGCGGTGTGCGG 4549
QY 909 AGCTTCTTTCGCTCGCATAGCTGTTCTTCTTCCGCGCGGCTGGTGTTCGCAAGGAT 968
Db 4548 CAGCTTTTTCGCGCATGATGCTGATGATGCTGCTGCGGCGGATGCGGAGCAGCAC 4489
QY 969 CTGGGATCTTTGACCGCGGTGCTGMAACAGCATGCTGTGAGCGCATTCGATGATGCG 1028
Db 4488 ATGGAGTGCCTGACGAGACTTTCATGGAACAGGTGCGCATGCGGTCTCATAGGCTGC 4429
QY 1029 GGTCTGCGCATGACGAGACTGCTCTTCTTGAACCTTTCACCGGACCGGTGTGAT 1088
Db 4428 CGGCTCGGCTGACGAGACCGGACCTTCTGCGTGTCTGACCGGAACCGGAGACA 4369
QY 1089 GGTGTGTTACATTTGGCTGCGACGCGCTGCTGCGAGTCAAGCTGCTTCCGTCATGG 1148
Db 4368 GCCCGCAATATCGGATCCCGCGGCGGCGTGAAGTGAACCTGTCCTTCCATGAG 4309
QY 1149 GAAATTTGAAGGCGTTCATAGTTCGCGACGATGAGAGCGCTACTGCGGTGCTCTGA 1208
Db 4308 CCGGTACGAATTAACGCTGAAGGCCCAACACACCGCGGTATCTGCGATAGCGCA 4249
QY 1209 ACAAATGCGCAAGGCTTTCAGAGGAGAGCTATTAAGTCTCGGTGATGCCATCAATT 1268
Db 4248 ATTTAGCGGCGGAGCTTTCAGAGGAGGTTCTATGAAATCGGGAATACGTCGAAATT 4189
QY 1269 GCGAATCTGCGCATCTCAAGAAAGTGTATGTTTGAAGCTGCAATTTGCTGAAGACTT 1328
Db 4188 TCCCGTCCGCGAGATCCCGCGCGGCTTTATTTTCATGATGCGCGCATGCGGAAACTT 4129
QY 1329 CAAGCTCTCTGAGGAGTATTTGTACGTTTGGGCAATTTGCGCAAGCGGCGGTTTGA 1388
Db 4128 CAAGCTGCAACCGGACCTGAGTGGCGGTTCGCGGCTTTGGCGGCAATTAATCAACT 4069
QY 1389 AGCGGCTTCTTACGCTGCTGAGAGTATGAGTGTGCTGCTCTGATGCTGAATGCTTGGATT 1448
Db 4068 GTTCCGCGGCTGATCTCGATGCTGATGCTCATATCGGCGGAAACCGGCGGAACTGGGCG 4009
```

```
QY 1449 GCTCGTGTTCGCGGCTCTTCGACTGCGCGCTTGTGCGGGCTAGAAAAGAGCGTC 1508
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4008 TCTGTCGTGTCCTTCATTACCGGCGCTACGTGAGCTTGTTCGCGTTGCAACATCTCTC 3949
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1509 GGAAGCCGAGGTGCTTGCAGTGAAGCCGATTTCGGGCTGTGTTGCTGACTGGCTCAACG 1568
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3948 CGATGCGGAGATCATCCGCGACCCGAGTGTCAAGGCGCAGATTGTGCGAAGCTTTCGGC 3889
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1569 ACTCAATCGAAGCAACTGGCAATGCCAATGCCATCATGTGGGTAGGGCTCCTCGATAC 1628
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3888 GCATCAAAAACAGCGCGATTCCGCAAGCCGCGTCATGCGCATTCGTGATGGAAGA 3829
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1629 GCGCGGTGATTAAGGCGAGTCACTGACAAAGGCTCGATCAACGAGCGCGCTGT 1688
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3828 CGCGCTGCGCTTCGAGAAAGCCAGGTTTACGACAAAGGATCGATCAACCAAGCGCGCGT 3769
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1689 TTGCAATGCGGCTGCGGCAAGTTGATGCGCTGTATCGTG 1729
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3768 GCTTGTGCATCGCAAGGAGCTCGTAGAGTCACCTTACGGCG 3728
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Search completed: February 7, 2005, 20:09:44
Job time : 7371 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2005, 23:26:45 ; Search time 44 Seconds
(without alignments)
887.757 Million cell updates/sec

Title: US-09-750-986D-30

Sequence: 1 MSLSEALLPPGRIERLERLH.....KVDAIYRGEDQSMIRDEATL 589

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Issued Patents AA:*

- 1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
- 2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
- 3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
- 4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
- 5: /cgn2_6/ptodata/1/1aa/PCITUS.COMB.pep:*
- 6: /cgn2_6/ptodata/1/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3082	100.0	589	US-08-976-063E-30	Sequence 30, Appl
2	1652.5	53.6	633	US-09-328-352-6519	Sequence 52, Appl
3	300	9.7	457	US-08-311-731A-52	Sequence 52, Appl
4	298	9.7	488	US-08-311-731A-283	Sequence 283, Appl
5	266	8.6	614	US-09-543-681A-7066	Sequence 7066, Appl
6	263.5	8.5	544	US-09-396-154-23	Sequence 23, Appl
7	263	8.5	634	US-09-418-963-3	Sequence 3, Appl
8	262.5	8.5	544	US-09-396-154-24	Sequence 24, Appl
9	262.5	8.5	544	US-09-396-154-45	Sequence 45, Appl
10	257.5	8.4	544	US-09-396-154-44	Sequence 44, Appl
11	255.5	8.3	544	US-09-396-154-21	Sequence 21, Appl
12	254.5	8.3	544	US-09-396-154-22	Sequence 22, Appl
13	252.5	8.2	544	US-09-396-154-19	Sequence 19, Appl
14	251.5	8.2	544	US-09-396-154-20	Sequence 20, Appl
15	249.5	8.1	550	US-09-602-628-2	Sequence 2, Appl
16	247.5	8.0	550	US-09-602-628-4	Sequence 4, Appl
17	246.5	8.0	550	US-08-354-240A-4	Sequence 4, Appl
18	246.5	8.0	550	US-08-354-240A-6	Sequence 6, Appl
19	246.5	8.0	550	US-08-718-425-5	Sequence 5, Appl
20	246.5	8.0	550	US-09-602-628-6	Sequence 6, Appl
21	246.5	8.0	550	US-09-602-628-8	Sequence 8, Appl
22	246.5	8.0	550	US-09-577-424-2	Sequence 2, Appl
23	244.5	7.9	616	US-10-140-002-206	Sequence 206, Appl
24	243.5	7.9	815	US-08-122-520C-9	Sequence 9, Appl
25	242.5	7.8	815	US-08-875-277A-4	Sequence 4, Appl
26	241.5	7.8	550	US-08-718-425-2	Sequence 2, Appl
27	240.5	7.8	549	US-08-354-240A-2	Sequence 2, Appl

28	240.5	7.8	550	3	US-08-867-352-23	Sequence 23, Appl
29	240.5	7.8	550	3	US-09-380-0618-21	Sequence 21, Appl
30	240.5	7.8	550	4	US-09-602-628-12	Sequence 12, Appl
31	240.5	7.8	550	4	US-09-396-154-31	Sequence 31, Appl
32	240.5	7.8	1242	4	US-09-488-270A-2	Sequence 2, Appl
33	237.5	7.7	550	3	US-08-875-277A-2	Sequence 2, Appl
34	237.5	7.7	550	3	US-08-487-183A-10	Sequence 10, Appl
35	237	7.7	739	4	US-10-140-002-478	Sequence 478, Appl
36	236.5	7.7	512	4	US-09-602-787A-86	Sequence 86, Appl
37	236.5	7.7	544	4	US-09-396-154-14	Sequence 14, Appl
38	236.5	7.7	568	4	US-09-602-787A-30	Sequence 30, Appl
39	234.5	7.6	544	4	US-09-396-154-17	Sequence 17, Appl
40	234	7.6	538	4	US-09-489-039A-13518	Sequence 13518, A
41	233	7.6	582	2	US-08-403-852D-20	Sequence 20, Appl
42	233	7.6	582	3	US-08-510-646B-21	Sequence 21, Appl
43	233	7.6	582	3	US-09-231-818-20	Sequence 20, Appl
44	233	7.6	582	4	US-09-635-359B-20	Sequence 20, Appl
45	232.5	7.5	550	3	US-09-380-0618-6	Sequence 6, Appl

ALIGNMENTS

```

RESULT 1
US-08-976-063E-30
; Sequence 30, Application US/08976063E
; Patent No. 6524831
; GENERAL INFORMATION:
; APPLICANT: Steinduchel, Alexander
; APPLICANT: Pfeifer, Horst
; TITLE OF INVENTION: SYNTHETIC ENZYMES FOR THE PRODUCTION OF CONIFERYL
; TITLE OF INVENTION: ALCOHOL, CONFERYLALDEHYDE, FERULIC ACID, VANILLIN AND
; FILE REFERENCE: VANILLIC ACID AND THEIR USE
; CURRENT APPLICATION NUMBER: US/08/976, 063E
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: 196 49 655.1 GERMANY
; PRIOR FILING DATE: 1996-11-29
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 30
; LENGTH: 589
; TYPE: PRT
; ORGANISM: not required under old rule
US-08-976-063E-30

Query Match      100.0%; Score 3082; DB 4; Length 589;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 589; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MRSLEALLPPGRIERLERLHMAKTRPEQTCVAAARANGEMRRISYAEFHNRAIAQSL 60
        1 MRSLEALLPPGRIERLERLHMAKTRPEQTCVAAARANGEMRRISYAEFHNRAIAQSL 60
DB      61 PYGSAERPLIYSGNDLEHLQAFGMYAGIYCPVSPAYSLSDLAUKRHYVGLQP 120
        61 PYGSAERPLIYSGNDLEHLQAFGMYAGIYCPVSPAYSLSDLAUKRHYVGLQP 120
QY      121 GLVEAADAAFFORAIETILPDVPAITRREGELAGRTVSEFDSLLEPGIEADNAFAATG 180
        121 GLVEAADAAFFORAIETILPDVPAITRREGELAGRTVSEFDSLLEPGIEADNAFAATG 180
DB      121 GLVPAADAAFFORAIETILPDVPAITRREGELAGRTVSEFDSLLEPGIEADNAFAATG 180
        121 GLVPAADAAFFORAIETILPDVPAITRREGELAGRTVSEFDSLLEPGIEADNAFAATG 180
QY      181 PDITAKLFTSGSTGLPKAVPTTORMICANQOMLQTFPVGGEPPVLVDMLPNNHTFGG 240
        181 PDITAKLFTSGSTGLPKAVPTTORMICANQOMLQTFPVGGEPPVLVDMLPNNHTFGG 240
DB      181 PDITAKLFTSGSTGLPKAVPTTORMICANQOMLQTFPVGGEPPVLVDMLPNNHTFGG 240
        181 PDITAKLFTSGSTGLPKAVPTTORMICANQOMLQTFPVGGEPPVLVDMLPNNHTFGG 240
QY      241 SHNIGIYLVNGGYLDDGKPTAGFAETLRNISEISPTAYLTVPKGMBELVGLERDST 300
        241 SHNIGIYLVNGGYLDDGKPTAGFAETLRNISEISPTAYLTVPKGMBELVGLERDST 300
DB      241 SHNIGIYLVNGGYLDDGKPTAGFAETLRNISEISPTAYLTVPKGMBELVGLERDST 300
        241 SHNIGIYLVNGGYLDDGKPTAGFAETLRNISEISPTAYLTVPKGMBELVGLERDST 300
QY      301 LRRFFARMKLFEEAAGLSQGIWDRLDRAVBOHCHGERIMMAGLMTETAPSCFTTGP 360
        301 LRRFFARMKLFEEAAGLSQGIWDRLDRAVBOHCHGERIMMAGLMTETAPSCFTTGP 360

```

Db 301 LRRERFFARMKLFEEFAAGLSQGIWDRLLDRVAEQHGERIRMMAGLGMTETAPSCFTTGP 360
 Qy 361 LSMAGYIGLPAPEGCEVLVVDGKLEGRPHGPHMSGYWRAPBONAOAFDEEGYCSGDA 420
 Db 361 LSMAGYIGLPAPEGCEVLVVDGKLEGRPHGPHMSGYWRAPBONAOAFDEEGYCSGDA 420
 Qy 421 IKLADPADPOKGLMFDGRIAEDFKLSGCVSVSGVPLRTAIVLEGSGYVLDVVVAAPDREC 480
 Db 421 IKLADPADPOKGLMFDGRIAEDFKLSGCVSVSGVPLRTAIVLEGSGYVLDVVVAAPDREC 480
 Qy 481 LGLVFPRLDLCALSGLEKESASPAETLASEPVRAMPADWLKRLNREATGNASIMVGL 540
 Db 481 LGLVFPRLDLCALSGLEKESASPAETLASEPVRAMPADWLKRLNREATGNASIMVGL 540
 Qy 541 LDTPEPIDKGEVTDKGSINORAVLQWRSAKYDALYRGEDQSMLEDEATL 589
 Db 541 LDTPEPIDKGEVTDKGSINORAVLQWRSAKYDALYRGEDQSMLEDEATL 589

RESULT 2

US-09-328-352-6519
 ; Sequence 6519, Application US/09328352
 ; Patent No. 6562958
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary L. Breton et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
 ; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: GTC99-03PA
 ; CURRENT APPLICATION NUMBER: US/09/328,352
 ; CURRENT FILING DATE: 1999-06-04
 ; NUMBER OF SEQ ID NOS: 8252
 ; SEQ ID NO 6519
 ; LENGTH: 633
 ; TYPE: PR1
 ; ORGANISM: Acinetobacter baumannii
 ; US-09-328-352-6519

Query Match 53.6%; Score 1652.5; DB 4; Length 633;
 Best Local Similarity 53.4%; Pred. No. 6.5e-168;
 Matches 310; Conservative 100; Mismatches 168; Indels 3; Gaps 3;

Qy 5 EALLPFGRLERLEHMAKTRPEOTCVAAAPANGEMRISYAEHFNVRALIAQSLPYGL 64
 Db 45 EOLKPYQKLTDRILHFAQTKPHIFAKRNAQOEHWKLSYAEVLEORAMHIAQALHARNL 104
 Qy 65 SAEPLIIVSGNDLEHQLAFGAMYAGIPYCPVSPAYSLSODLAKLRHIVGLQPLVLF 124
 Db 105 SGERPLVILSGNDLEHQLAFGAMYAGIPYCPVSPAYSLSODLAKLRHIVGLQPLVLF 164
 Qy 125 AADAAPFORAIEITLLPDDVPAIFTRGELAGRRTVSPDLSLEOPGCIADNAFAATGPDJI 184
 Db 165 ASDGQAFAKAIQACITPDIEVNTKGIYGDQICTSPGSLDTPVS-NVGEFYQTLDBNOI 223
 Qy 185 AKLEFTSGSKTKRAVATTQRMICANQOMLLQTFPVGGEPPVLVMDLPMNHTFGGSHNT 244
 Db 224 AKLEFTSGSKTKRAVATTQRMICANQOMLLQTFPVGGEPPVLVMDLPMNHTFGGSHNT 283
 Qy 245 GIVLYNGGTYYLDGKPTAOGFAETLNLSEISPTALVTPKGEELVGLERDSTLRER 304
 Db 284 GIALYNGGTYYLDGKPTAOGFAETLNLSEISPTALVTPKGEELVGLERDSTLRER 343
 Qy 305 FFARMKLFEEFAAGLSQGIWDRLLDRVAEQHGERIRMMAGLGMTETAPSCFTTGP 364
 Db 344 FFARMKLFEEFAAGLSQGIWDRLLDRVAEQHGERIRMMAGLGMTETAPSCFTTGP 403
 Qy 365 GYIGLPAPEGCEVLVVDGKLEGRPHGPHMSGYWRAPBONAOAFDEEGYCSGDA 423
 Db 404 GYIGLPAPEGCEVLVVDGKLEGRPHGPHMSGYWRAPBONAOAFDEEGYCSGDA 463
 Qy 424 ADADADPOKGLMFDGRIAEDFKLSGCVSVSGVPLRTAIVLEGSGYVLDVVVAAPDREC 483
 Db 464 ADADADPOKGLMFDGRIAEDFKLSGCVSVSGVPLRTAIVLEGSGYVLDVVVAAPDREC 523

Qy 484 LVPEPRLDLCALSGLEKESASPAETLASEPVRAMPADWLKRLNREATGNASIMVGL 542
 Db 524 LVPEPRLDLCALSGLEKESASPAETLASEPVRAMPADWLKRLNREATGNASIMVGL 583
 Qy 543 TPEPIDKGEVTDKGSINORAVLQWRSAKYDALYRGEDQSMLEDEATL 589
 Db 584 TPEPIDKGEVTDKGSINORAVLQWRSAKYDALYRGEDQSMLEDEATL 624

RESULT 3

US-08-311-731A-52
 ; Sequence 52, Application US/08311731A
 ; Patent No. 6583266
 ; GENERAL INFORMATION:
 ; APPLICANT: SMITH, DOUGLAS
 ; APPLICANT: MAO, JEN-I
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
 ; TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR
 ; NUMBER OF SEQUENCES: 411
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
 ; STREET: 600 ATLANTIC AVENUE
 ; CITY: BOSTON
 ; STATE: MASSACHUSETTS
 ; COUNTRY: USA
 ; ZIP: 02210
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/311,731A
 ; FILING DATE:
 ; CLASSIFICATION: 530
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: GATES, EDWARD R.
 ; REGISTRATION NUMBER: 31,616
 ; REFERENCE/DOCKET NUMBER: C0044/7125
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 617/720-3500
 ; TELEFAX: 617/720-2441
 ; INFORMATION FOR SEQ ID NO: 52:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 457 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; HYPOTHEICAL: YES
 ; ORIGINAL SOURCE:
 ; ORGANISM: MYCOBACTERIUM LEPRAE
 ; US-08-311-731A-52

Query Match 9.7%; Score 300; DB 4; Length 457;
 Best Local Similarity 26.2%; Pred. No. 5.5e-23;
 Matches 143; Conservative 67; Mismatches 210; Indels 126; Gaps 27;

Qy 44 SVAEHFNRAIQSLPYGLSAERLLIVSGNDLEHQLAFGAMYAGIPYCPVSPAYSL 103
 Db 13 SVAEHFNRAIQSLPYGLSAERLLIVSGNDLEHQLAFGAMYAGIPYCPVSPAYSL 157
 Qy 104 LSODLAKLRHIV-----GLQPLVPAADAPFORAIEITLLPDDVPAIFTRGELAGRRT 157
 Db 58 VER-----RHMLDSGAQAMLGPGP--SADSA-----PDGLPHI-----P 90
 Qy 158 VSPDLSLBOPGCIADNAPATGPDITIAKFLFTSGSTKLPAVPTTQRMICANQOMLLQ 217
 Db 91 VOLDA-----RSWNRYPESPDPDATAMVITSGTGPCKGLVLSRRAIVDLDAQA 142
 Qy 218 FVPGEPPVLVMDLPMNHTFG-----GSHNIGIVLYNGGTYYLDGKPTAOGFAETLR 271
 Db 143 WQNTAAD--VLVHGLPLFHVHGLVGLGLSLRI-----GNRFVHTGKPTTAIVQA-- 191

QY 272 NLSEISPTAYLVTPKGEELVGALEBDSTLEREFPAAMKLPFPAAGLSGIMDRDRA 331
 DB 192 -CSBAGSLYFGVPTVWSRLVA---DEAV-ARALRPARLLVSGASLPVPVFDRLAHT 245
 QY 332 EOHGGERIRRMAGLGMTETAPSCFTTGPLSMAGYIGLPAPGCEVKL-----VPVDGK 384
 DB 246 GHRPIER-----YGSTESLITLSTLADGERBAGWGLPLAGVOTRLVDSGGCVPYDGE 299
 QY 385 LEGRF--HGPHVSGYKRAPEQNAQAFDEEGYICSGDAIKLADPADPQKGLMPDRIAD 442
 DB 300 TVRQLQVRSPTMGYLNREAPATAEAFDEDEGWRTGD-VAVVDSGGMH--IVGRESVD 355
 QY 443 FKLSGVFVSVPLRTAVLEGGSYLVV-VVAAPDEECGLVFPRLDORALSGLGE 501
 DB 356 LKLGGRIGAGEL--EMALLGHPDVREAVAVGLPD-EDLQGRIVAFV-----VGAE 404
 QY 502 ASDAEVLASBPVAMFADMLKRLNREATGNASRIWVGLDTPPSIDKEVTDKSGINOR 561
 DB 405 ALDADELIN-----YVAQQLSIHKRPRE-----VRFDALPA-TRWERSKSSCSSR 450
 QY 562 AVLQWR 567
 DB 451 A--DMR 454

RESULT 4
 US-08-311-731A-283
 ; Sequence 283, Application US/08311731A
 ; Patent No. 6583266
 ; GENERAL INFORMATION:
 ; APPLICANT: SMITH, DOUGLAS
 ; APPLICANT: MAO, JEN-1
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
 ; TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR
 ; NUMBER OF SEQUENCES: 411
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
 ; STREET: 600 ATLANTIC AVENUE
 ; CITY: BOSTON
 ; STATE: MASSACHUSETTS
 ; COUNTRY: USA
 ; ZIP: 02210
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/311,731A
 ; FILING DATE:
 ; CLASSIFICATION: 530
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: GATES, EDWARD R.
 ; REGISTRATION NUMBER: 31,616
 ; REFERENCE/DOCKET NUMBER: C0044/7125
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 617/720-3500
 ; TELEFAX: 617/720-2441
 ; INFORMATION FOR SEQ ID NO: 283:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 488 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; HYPOTHEICAL: YES
 ; ORIGINAL SOURCE:
 ; ORGANISM: Mycobacterium leprae
 ; US-08-311-731A-283
 Query Match 9.7%; Score 298; DB 4; Length 488;
 Best Local Similarity 26.4%; Pred. No. 1e-22;

Matches 135; Conservative 61; Mismatches 196; Indels 120; Gaps 24;
 QY 44 SYAEMFNPAIAQSLPFLYLSAERPLLYSGNDLHQLAFAAMVAGIYCVSPAYSL 103
 DB 50 SVABRVGAGLVAVLATP--TASTVLATP-----CCLLNGVAVVPADIGV 94
 QY 104 LSODLAKLRHIV-----GLLOGLVFAADAPFORALETILPDDVPAIFTRGELARRT 157
 DB 95 VER-----RMLVDSGAQAWLGRGP--SADSA-----FDGLPHI-----P 127
 QY 158 VSFDSLLEOPGILEADNFAATPDTIAKFLFTSGSTKLKPAVPTTQMLCANQOMLLOT 217
 DB 128 VQIDA-----RSMNRYPEPSPDPAIAWITYSGTTGPKPKVILSRRAIAVLDALAAQ 179
 QY 218 FVFGSEEPVLYNMLPNNHTFG-----GSHNIGIYLVNGTYVLDGKRTAGPAETLR 271
 DB 180 WQMTAAD--VLVHGLPLFHHGLVLGLLSGLR-----GNRPVHTGKPTPAVQA-- 228
 QY 272 NLSEISPTAYLVTPKGEELVGALEBDSTLEREFPAAMKLPFPAAGLSGIMDRDRA 331
 DB 229 -CSBAGSLYFGVPTVWSRLVA---DEAV-ARALRPARLLVSGASLPVPVFDRLAHT 282
 QY 332 EOHGGERIRRMAGLGMTETAPSCFTTGPLSMAGYIGLPAPGCEVKL-----VPVDGK 384
 DB 283 GHRPIER-----YGSTESLITLSTLADGERBAGWGLPLAGVOTRLVDSGGCVPYDGE 336
 QY 385 LEGRF--HGPHVSGYKRAPEQNAQAFDEEGYICSGDAIKLADPADPQKGLMPDRIAD 442
 DB 337 TVRQLQVRSPTMGYLNREAPATAEAFDEDEGWRTGD-VAVVDSGGMH--IVGRESVD 392
 QY 443 FKLSGVFVSVPLRTAVLEGGSYLVV-VVAAPDEECGLVFPRLDORALSGLGE 501
 DB 393 LKLGGRIGAGEL--EMALLGHPDVREAVAVGLPD-EDLQGRIVAFV-----VGAE 441
 QY 502 ASDAEVLAS-----EPVAMFADMLK 523
 DB 442 ALDADELINYVAQQLSIHKRPREVRFDALPR 473

RESULT 5
 US-09-543-681A-7066
 ; Sequence 7066, Application US/09543681A
 ; Patent No. 6605709
 ; GENERAL INFORMATION:
 ; APPLICANT: GARY BRETTON
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
 ; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 2709, 1002-001
 ; CURRENT APPLICATION NUMBER: US/09/543,681A
 ; CURRENT FILING DATE: 2000-04-05
 ; PRIOR APPLICATION NUMBER: US 60/128,706
 ; PRIOR FILING DATE: 1999-04-09
 ; NUMBER OF SEQ ID NOS: 8344
 ; SEQ ID NO 7066
 ; LENGTH: 614
 ; TYPE: PRT
 ; ORGANISM: Proteus mirabilis
 ; US-09-543-681A-7066
 Query Match 8.6%; Score 266; DB 4; Length 614;
 Best Local Similarity 21.6%; Pred. No. 4e-19;
 Matches 140; Conservative 105; Mismatches 270; Indels 132; Gaps 23;
 QY 14 ILERLEHMAKTRBQCVAAAPANGERRISYVEMFNPAIAQSLPFLYLSAERPLLY 73
 DB 24 ILIRLQWQFSSSGKTYRQWEAKSB-IAMSVASAKTRALNALDMGVAQVENVGIF 82
 QY 74 SGN-----DLHLQAFAGAMVAGIYCVSPAYSLSDLKLRIHIVGLQGLVFAA 126
 DB 83 SONSIDMSIADIATLQLR-----AVTVPIYATSSVEQA--AYIIDADIRIIFVG 130
 QY 127 DAAPFGRAIR-----TLPDDVPAIFTRGELARRTVSFDLSLEOPG 169

Db 131 DOKEYDVSESLALLCPQLKHHIVFNSQVNLNPTIPSYHLTNLISDCQS -OYDSLLQO--- 186
Qy 170 IEADNFAATGPPTIAKFLFTSGSTKLPKAVPTTQRMLCANOQMLQTFPVGSEBPVLV 229
Db 187 -RIDECL-----DDLFLITTSSTGTGEPKGMLDYTNLAS--QYIHLQKLSISDKDVL 239
Qy 230 DMLPMNHTFGSHNIGIVLNGSTYYLDDKPTAOCFAETLRNLSEISPTAYLTVPKGWE 289
Db 240 CFLPLSHVFERRAWSFYVMHTGALINVYLTDTHAVREANA-----AVKPTVMCAVPRFYE 292
Qy 290 ELVGALEBDSLTAEERFAPKULFFFA-----AAGLSOGI---W-----DRDDR 329
Db 293 KYVSAIOEKVSOAPRL--RQMFKMLKQGEKORQAHLARSLGVSQWCFYRPAKKVLQ 350
Qy 330 VAEQHCGERIRMA-----GLQMTETAPSCFTTTPGLSMAGY 366
Db 351 PLRQIIGGRVRFPAAGARLDVAIGFPLAAGINIKKYGMTETCATVSCWENHYLGGS 410
Qy 367 IGLPAPGCEVYLVVDKLEGRFHPVMSGYWRAPQNAQAFDEBGYYCGDAIKLADP 426
Db 411 IGTPLPGVEYR---IGBENEIOVRGSIWMKGYFNKPEETVAAPTEQDQMLRTGAGAL--- 464
Qy 427 ADPQKGMFPGRIAEADPKLSSGVFVSVPRTRAVLDEGSYVDVYVAADRECLGLLVF 466
Db 465 -DEDGMLYITERLKDLMKTSNGKYIA--PQMIEGTLGQDFRIFHIAVIAPTRRFVSALIV 521
Qy 487 PRLLDC-----RALSGKEASDAEVLASEPVRAWPAFADMLKRLREATGNASRIWV 538
Db 522 P----CFDMLSEARALN--LKYHDRLEHLRHKIKELFERLRERQNRN--FASFHQYKRF 574
Qy 539 GLIDTPPSIDKGEVTDKGSINQRAVLQWRSKAVDALYRGEDQMLRD 585
Db 575 TLAEFTWESGELT-----PTLKLRRKTIISERYRNEIQWYQE 613

RESULT 6
US-09-396-154-23
; Sequence 23, Application US/09396154
; Patent No. 6602677
; GENERAL INFORMATION:
; APPLICANT: Wood, Keith V.
; APPLICANT: Hall, Mary P.
; TITLE OF INVENTION: Thermostable luciferases and methods of
; FILE REFERENCE: 341.012US1
; CURRENT APPLICATION NUMBER: US/09/396,154
; EARLIER FILING DATE: 1998-09-15
; EARLIER APPLICATION NUMBER: US 09/156,946
; EARLIER FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: PCT/US98/19494
; EARLIER FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: US 60/059,379
; EARLIER FILING DATE: 1997-09-19
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 23
; LENGTH: 544
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Mutant luciferase
; NAME/KEY: UNSURE
; LOCATION: (354)...(355)
; OTHER INFORMATION: Unknown amino acids
US-09-396-154-23

Query Match 8.5%; Score 263.5; DB 4; Length 544;
Best Local Similarity 22.5%; Pred. No. 6,1e-19;
Matches 130; Conservative 86; Mismatches 226; Indels 137; Gaps 24;
Qy 30 CVAAPANGWRKRISIAEMHNVRALAQSLPYGLSARPLLVSGNDLEHLQALFQAMV 89

Db 38 CIALTNAAHK-ENVLYEEFLKLSCLRAESFKYKGLKQNDTIAVCSNGQLFPLPIASLY 96
Qy 90 AGIPYCPVSPAYSLSDQLAKLHHIVGLQPGLVFAPADAAPFORALETLLPDDVPAIFTR 149
Db 97 LGIIVAPVNDKY--IRRELI---HSGIYKPRRIIFCSKRT--PQKVLN-----VK 139
Qy 150 GELAGRRIVSPSLLEQPGIEADNAPAAATG-----BDTAKFLFTSG 192
Db 140 SKLSVETIILIDLNEDLGQCQANFNISQNDINDVKKKFSPNRDQVALIMFSSG 199
Qy 193 STLKPAVPTTQMLCANQOMLQTFPVG---EEEPVLVDMLPMNHTFGSHNIG---- 245
Db 200 TGLPFGVNLTHKNIVA--RPSLAKDPTFGNAINPTTALITVLIFRHHGGMATTLGYFTC 257
Qy 246 ---IVLNGSTYYLDDKGPFAQFAET--LRNLSEISPTAYLTVPGMEELVGALEBRDT 300
Db 258 GRRVULMH-----TFEEKLTLQSIQDVKVESTLLV---TLMAFLAKSAL 299
Qy 301 LRERFAPKMLFPFAAGLSOGIMDRLDRAVEQHCGERIR-----MMAGLQMTETAPB 353
Db 300 VEKYDSLHKEIASGGAPLSKEI-----GEMVKKRFLNFRQGYGLTETTS 347
Qy 354 CTFTTGPLSMAGYIGLPAACEVYLV-PYDGKLEG-----RFHGPVMSGYWRAPQNR 405
Db 348 VLIPTPKXAPPGSGTKIVFHAQKVDPPTGKILGNEBEGELYFKGPMIMKGYNNBEAT 407
Qy 406 AQAFDEGYYCGSDAIKLADPADPQKLMFDDGRIADFKLSGCVFVSVPPLRTRAVLLEG 465
Db 408 KAIIDNDGMLRSGDI-----AYYD-----NDGHFYIVRLSLIKYKG- 445
Qy 446 SYVLADVVAAPRECLGLLVFPRLLDCRALSGKEASDAEVLASEPVRAWPAFADMLKRLN 525
Db 446 ----YQVAPAEIEGI-LIQHPYIVDA-GVTGIPDEAAGELPAAQVAVVQTG-----KYLN 493
Qy 526 RE-----ATGNASRIWV-----VGLIDTPPSIDKGEVTDK 555
Db 494 EQIVQDVSSQVSTAKMLRGVAFLEDEIFKSGTGKIDRK 532

RESULT 7
US-09-418-963-3
; Sequence 3, Application US/09418963
; Patent No. 6664039
; GENERAL INFORMATION:
; APPLICANT: Min, Kyung-Tai
; APPLICANT: Benzer, Seymour
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR MODULATING
; FILE REFERENCE: 06618-367001
; CURRENT APPLICATION NUMBER: US/09/418,963
; EARLIER FILING DATE: 1997-10-14
; PRIOR APPLICATION NUMBER: US 60/104,298
; PRIOR FILING DATE: 1998-10-14
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 634
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-418-963-3

Query Match 8.5%; Score 263; DB 4; Length 634;
Best Local Similarity 21.4%; Pred. No. 8.9e-19;
Matches 135; Conservative 93; Mismatches 282; Indels 120; Gaps 20;
Qy 39 EWRRIYAEHFHNVRALAQSLPYGLSARPLLVSGNDLEHLQALFQAMVAG---IPY 94
Db 39 KWEHISYSQYVLLARBAKPGFLKLGKQASVALILGFNSPEWFFSAGVGVFAGIVTGIY 98
Qy 95 CPVSP-AYSLSODLAKLHHIVGLQPGLVFADAPAPORAITFTI----- 139
Db 99 TTSSPACQYIATDC-----ANVIWDT---OKOLEKILKIMQPLKALAVI 144


```

Oy 140 -----PDDVPAIFETGSGELAGRTVSPDLSLEQPGGIEADNAPAAAGPDITIAFLFSGST 194
Db 145 YKEPPNKANAYITMBE-----FMELGNEVPEBALDITDQOPNCCVLYVTSIT 196
Oy 195 KLPKAVPPTTORMLCANQOMLDQTFPVFGEE--PVLVDMLPMNHTFGSGSHNIGIYLYNG 252
Db 197 GNPKVMLSQDNITWTATARYGAGGIRAPAEVQGEVVVSYLPLSHLAQIYDLMTGIGMKA 256
Oy 253 TYVLDGKPTAGFAETLRNLSEISPTAYLYVPKGMEILVGLERDSTLERFPFAMKLF 312
Db 257 QVCFAPEDALKGLVNTLR---EVEPTSHMGVPRWKEKIMERIQ-EVAQSGFIRKMLL 312
Oy 313 FFAAAGLSQGT-----WDLRVAE-QH- 334
Db 313 WAMSYTLLEQNLTICPGSDKLPFTTRLADLYLAKVQALGPAKCKNPFYAAPMAETQHF 372
Oy 335 -GGERIRMMAGLGMETAPSCFTTGPISMAGY-IGLPARGCEVLYVPDGLKEGR--PH 390
Db 373 FLGLNIRLYAGYGLSETS-GPHFMSFPYRYLYSGSKLVPGKRVKLVNDADGIGICIM 431
Oy 391 GPHVWGYRPAPEQNAQAFDEBGYYCSGDALKIADPADPQKGLMPDGRIAEDFKLSGVF 450
Db 432 GRTIFMGVGLNMEDKCEALDEEGMLHTGBAGRL----DADGFLYITGRKELLITRAGSN 487
Oy 451 VSGVPLRFRVALEGGSYVLDVVAAADRECLGLVF-----PRLD-----C 492
Db 488 VPPVPIEEAVKME-LPIISNAMLIDQRKFSLMLTLKCTLPDTPSDQTDNLTQEAVERC 546
Oy 493 RALSLGEGASDAEVLASEPVAWADMLKRLNREATGNASHIMWGILDTPPSIDKGEV 552
Db 547 QRVGSRATTVSIIIEKKDEAVYQALIEGIRRVNMMAAPPHYIQKVALIERDPSISGGEL 606
Oy 553 TDKGSINORAVLQWRSKADVALYRGEDQSM 582
Db 607 GPTMKLKRPLYLEKKGITDSFY--QEQKM 634

RESULT 8
US-09-396-154-24
; Sequence 24, Application US/09396154
; Patent No. 6602677
; GENERAL INFORMATION:
; APPLICANT: Wood, Keith V.
; APPLICANT: Hall, Mary P.
; TITLE OF INVENTION: Thermostable luciferases and methods of
; FILE OF INVENTION: production
; FILE REFERENCE: 341.012U1
; CURRENT APPLICATION NUMBER: US/09/396,154
; CURRENT FILING DATE: 1999-09-15
; EARLIER APPLICATION NUMBER: US 09/156,946
; EARLIER FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: PCT/US98/19494
; EARLIER FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: US 60/059,379
; EARLIER FILING DATE: 1997-09-19
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 24
; LENGTH: 544
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Mutant luciferase
US-09-396-154-24

```

Query Match Similarity: 8.5%; Score 262.5; DB 4; length 544;
Beat Local Similarity: 22.6%; Pred. No. 7,8e-19;
Matches 131; conservative 85; mismatches 222; indels 137; gaps 22;

Cy 30 CVAARANGKRRISYAEFHNVRALQSLPYGISAERPLIYSGNDLEHQLAFAMCY 89
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 38 CIATLNATK - ENVLYEEFLTKSCHLASSFKRYGKKDNTAVCGENGLQFFLPVIALY 96

```

QY 90 AGIPCPSPAAVSLLSODLAKRHTIVGLGCPGVFAADAAPFORALLETILDPDVPAIFTR 149
Db 97 LGIYAPVNDKY--IERELI--HSLGIVKPIRVFOSKNT--FOKVLN-----YK 139
QY 150 GELAGRIVSPDSLLOPGGIEADNAPAFATGPDY-----IAKPLFTSG 192
Db 140 SKLSKIEIETIIILDLNEDLGGYOCNLNFISSQNSDNLDYKKFKRPFSPNRDDQVALIMFSSG 199
QY 193 STKLPRKAVPTTQRMLCANOQMLQTFPVFG--EEPRVLVDMLPMNHTFGSHNIG--- 245
Db 200 TTGLPKGVMLTHKNIVA--RFSIADLPFGNALINPTALTITVYIPEHHGGMHTTLGFTC 257
QY 246 ---IYLVNGGYIYLDGKPTAGQAFET--LRNLSEISPAIYLTVPKGMSELYGALERDST 300
Db 258 GFRVVLNM-----TFEKKFLQSLQDYKVESTILLP---TLMAFLAKSL 299
QY 301 LRERPFARMKLFFAAGISGIGWRLDRVABQHCGERIR-----MMAGLGMEPTAS 353
Db 300 VEKYLDSHUKELASGCAPLSKET-----GENVKKRPKLMFVNOGIGLITTTSA 347
QY 354 CTFITGPLSMAGYIGLPAFGCEVKLV-PVDCKLEB-----RFGHPHVMSGYMARPEON 405
Db 348 VLITPKGDAKPGESTKIVPRHAVKVYDDPTTGKILLENBEGELYFGKPMIMKGYUNNEAXT 407
QY 406 AOAPEDEGYSCGDAIKLADPADPQKGIMFDGRIAEDFKLSSGVFVSPGLPRTRAVLBEG 465
Db 408 KAIINDGMILRSGLD-----AYYD-----NGHFYIVRLRSLLIKYKG- 445
QY 466 SYVLDDVVAAPRECLGILLVPRLLDCLRALSGLSKEASPAEVLASPRVAPAMFADWLKRLN 525
Db 446 -----YOVAPEAIEEI-LDQHPYIVDA-GVTBIPDEAAELPAAAGVVOYTG-----KYLN 453
QY 526 RE-----ATGNASRIMW-----VGLLDTSPSIDKGEVTDK 555
Db 494 EOIVODIYVASQVSTAKMLRGVYKFLDELIPKSGTGKIDRK 532

```

```

US-09-396-154-45
; RESULT 9
; Sequence 45, Application US/09396154
; Patent No. 6602677
; GENERAL INFORMATION:
; APPLICANT: Wood, Keith V.
; APPLICANT: Hall, Mary P.
; TITLE OF INVENTION: Thermostable luciferases and methods of
; FILE OF INVENTION: production
; FILE REFERENCE: 341.012US1
; CURRENT APPLICATION NUMBER: US/09/396,154
; CURRENT FILING DATE: 1999-09-15
; EARLIER APPLICATION NUMBER: US 09/566,946
; EARLIER FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: PCT/US98/19494
; EARLIER FILING DATE: 1998-09-18
; EARLIER FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: US 60/059,379
; EARLIER FILING DATE: 1997-09-19
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 45
; LENGTH: 544
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Mutant luciferase
US-09-396-154-45

```

```
Query Match 8.5%; Score 262.5; DB 4; Length 544;  
Best Local Similarity 22.5%; Pred. No. 7.9e-19;  
Matches 130; Conservative 86; Mismatches 226; Indels 137; Gaps 24;
```

[illegible]

Query Match	8.4%;	Score 257.5;	DB 4;	Length 544;
Best Local Similarity	21.9%;	Pred. No. 2.7e-18;		
Matches 124;	Conservative 83;	Mismatches 227;	Indels 131;	Gaps 22;

```

Qy      30  CVAARANGEMRISYEMFHNVAIOSLLPYGISAEPLLIYSGNDLEHLQIACGAMY 89
          | | : : | | : : | | : : | | : : |
Db      38  CIALTNAHTK-ENVLYEEFKLSCLAESEFKYTGKQNDTIVCSENSLQFFLPVITASLY 96

```

```

QY 90 AGIPCEPBPAASSLISODIAKRNHVGLJGVLGFAADAPFORAITEILPDDVPAIFTR 149
Db 97 LGIIVAPVNDKI--IERELI---HSLGVRKRIYFCGKNT--FQKVLN-----VK 139

QY 150 GELAGRIYSPDSLLEQPGGIEADNAPAAATGPD-----IAKLFITSG 192
Db 140 SKKSISETIIIIILDNDIDGGYOCUNLPFSIONSDNLDVKRKPYSFNRDOVALIMFSSG 199

QY 193 STKLPIAYVTTQORMCANQOMLOTPEPVG---BEPPVLVDMLEPMNTHFGCSHNIG--- 245
Db 200 TTSLPBGVWLTNKNIVA--RISIADPRTFGNAINPTAIIILVIRPHHGFQGMNTLIGFTC 257

QY 246 ---IIVNGGTYYLDDGKPTAGPAET--IENLSEISPTAYLIVPKGMBELVGLAEBDST 300
Db 258 GFRVVLVHM-----TFEEKLFLOSLODYKVESTILVP---TLMFLAKSAL 299

QY 301 LBERFPAPAKULFPFAAGLSGSGIMWRDLRVAHQCGERIR-----MMAGLGMTETAPS 353
Db 300 VEKIDYLSHKETAISGAPLSKEI-----GEMVKRRFPLANTVVRQGYGLTETSA 347

QY 354 CFTTGTPLSMAGYIGLAPGCEVKLV--EVDGKLEB-----RPHGHVMSGYWRABEON 405
Db 348 VLIITPKGADPSTGTGIVPFHAVKRVADPTTGKILGPNBPELYKKGPMINKMGYNEEAT 407

QY 406 AQAFDEEGYVCGGDALIKLADPADPQKGLMFGRIIAEDFKLSSG---FVSGVPLRTAYL 462
Db 408 KAIINDDGLWRSGD--IAYYDN-----DGHFIYDRLSLLKYKGYOAPAPAEIBGIL 457

QY 463 EGGSYVLDDVVAAPDPRECTGLIVFPRLIDCRALSGLGEASDAEVLASEPVRAMFAD--- 519
Db 458 LGHPYIVDAGVTGIDPEAAG-----ELPAGAVVQGTGKYLNEQIVODVYASQVS 506

QY 520 ---WLK-----RLNREATGNASR 534
Db 507 TAWMLRGVYIFLDEIIPKSGTGKIDR 531

```

```

RESULT 11
US-09-396-154-21
; Sequence 21, Application US/09396154
; Patent No. 6602677
; GENERAL INFORMATION:
; APPLICANT: Wood, Keith V.
; APPLICANT: Hall, Mary P.
; TITLE OF INVENTION: Thermostable luciferases and methods of
; title of invention: production
; FILE REFERENCE: 341.012US1
; CURRENT APPLICATION NUMBER: US/09/396,154
; CURRENT FILING DATE: 1999-09-15
; EARLIER APPLICATION NUMBER: US 09/156,946
; EARLIER FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: PCT/US98/19494
; EARLIER FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: US 60/059,379
; EARLIER FILING DATE: 1997-09-19
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 21
; LENGTH: 544
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Mutant Luciferase
; NAME/KEY: UNSURE
; LOCATION: (354)...(355)
; OTHER INFORMATION: Unknown amino acids
; US-09-396-154-21

```

Query Match	8.3%	Score	255.5	DB	4	length	544
Best Local Similarity	22.9%	Pred. No.	4.4e-18				
Matches 132, Conservative	92	Mismatches	220	Indels	133	Gaps	27

Query Match 8.3%; Score 254.5; DB 4; Length 544;
 Best Local Similarity 22.5%; Pred. No.5.6e-19;
 Match 130; Conservative 87; Mismatches 225; Indels 137; Gaps 24;

```

QY      30  CVAARAANGEMRRISYAEMFHNVAIOSSLVYGLSABRPLLIVSGNDLEHLOAFGAMY 89
      38  CIALTNAMHTK-ENVLYEBFLKLSGRLASBFRKYLKQNDITAVCSENGLOPFLEVIASLY 96
QY      90  AGIPYCPSPAYSLLSODDLAKRLIVGLQDGLVFAADAAFPQAIETILPDDVPAIFTR 149
      97  LGIIAAPYSDRY--VERBELI---HSLGIYKRRIIFCSKNT-FQKVLN-----VK 139
QY      150  GELAGRRVSPDSLLDEPGGIEADNMAAATGPT-----IAKFLFTSG 192
      140  SKLKYVEITILLDNLDELGYQCLANNFTSONSDSNLDVKKFKPNSFYKDDOVALVMTSSG 199
QY      193  STKLPAKVPPTTQRMLCANQOMLQTFPVFG---EEPPVLVDMLEWNTTFCGSHNIG--- 245
      200  TTGVPKGVMLTHKNIYA--RPSLAKDPTFGNAINPTAIIIVLPIFHFGFMMTLGYFTC 257
QY      246  ---IVLYNGGYIYLDGKPTPAQGAET--LRNLSEIPTAYLYPYKMEELVGLAENDST 300
      258  GFRVLYNH-----TTEKKLFLOSLODYKVESTLVP---TLMAFLASAL 299
QY      301  LREPFAMKLEFPAAAGLSGQIWDRLDRAVBOHCGERIR-----MMAGLGWTETAPS 353
      300  VEKIDLSHUKIASGAPLSKEI-----GEMVKKRFKLTINVROGYLTETTTSA 347
QY      354  CTFTTGPLSMAGYIGLPAFGCEVLY-V-PVDKLEG-----RPHGPHVMSGYRAPDON 405
      348  VLIPTPKXXARPGSTGKIVPFAVAVKVPDPTTGKILGPNMETGELYFKGAMIMKGYNNBEAT 407
QY      406  AQADDEBEYYSSGAITKLADPADQOKGLMPDGRLAEPFKLSSGVFVGLPRTAAVLEGG 465
      408  KAIDKQGWLRSGOI-----AYYD-----NDGHFYIVDRKSLIKYKG- 445
QY      466  SYVDLVAAADRECLGLVPEPLLDORALSGLSEKASDAEVLSEPVAFADMLKRLN 525
      446  -----YQVAPEIETIGI-LLOHPYIVDA-GYTGIDDEAAGELPAAGVYVQTC-----KTLN 493
QY      526  REATGN-----ASRIMW-----VGLLDTPPSIDKGEVTDK 555
      494  EQIVQNFVSSQVSTAKMLRGGVKFLDEIPKSGSTGKIDRK 532

RESULT 13
US-09-396-154-19
; Sequence 19, Application US/09396154
; Patent No. 6602677
; GENERAL INFORMATION:
; APPLICANT: Wood, Keith V.
; APPLICANT: Hall, Mary P.
; TITLE OF INVENTION: Thermostable luciferases and methods of
; FILE REFERENCE: 341.012US1
; CURRENT APPLICATION NUMBER: US/09/396,154
; CURRENT FILING DATE: 1999-09-15
; EARLIER APPLICATION NUMBER: US 09/156,946
; EARLIER FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: PCT/US98/19494
; EARLIER FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: US 60/059,379
; EARLIER FILING DATE: 1997-09-19
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 19
; LENGTH: 544
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Mutant luciferase
US-09-396-154-19

```

Query Match 8.2%; Score 252.5; DB 4; Length 544;
Best Local Similarity 22.3%; Pred. No. 9.2e-18;
Matches 129; Conservative 87; Mismatches 226; Indels 137; Gaps 24;

```
QY 30 CVAARANGEMRISYAEMFHVRAIAQSLPFGLSAEPRLLVSGNDLEHLQAFGAMY 89
DB 38 CIALTNHNR-ENVLYEEFLKSCRLAESFKKYGKLNQNTIAVCSENGLOFLPVIASLY 96
QY 90 AGIPYCPVSPAYSLSDILAKLRIHVGLQPLVFADAAAPFORAIFTLIPDVPAIFTR 149
DB 97 LGIIAAPVSDKY--IERELI---HSLGIKPRILIFCSKNT-FQKVLN-----VK 139
QY 150 GELAGRTVSPDLSLEPGIIEADNAPATGPDT-----IAKFLFTSG 192
DB 140 SKLSKSVETIILDLNEDLGVCQCLNNFISQNSDSNLDVKKFKPYSFNRDQVALVMESSG 199
QY 193 STLPRVPTTQMLCANQOMLQTPFVFG---EEPRVLDVLPVMTTFGSHNIG---- 245
DB 200 TTQVPKGVMLTHNIVA--RFSIAKQDPTFGNAINPTTALITVLPFHGFGMMTTLGFTG 257
QY 246 ---IVLYNGSTYYLDGKPTAOGFAET--LRNLSEISPTAYLTVPKGMELVGALEERDST 300
DB 258 GFRVVLNH-----TFEEKLFLQSLQDYKVESTLVP-----TLMAFLAKSAL 299
QY 301 LREPRFAPRMKLFPPAAAGLSQGIWDRIDRVABQHCGRIR-----MMAGLMTEPTAPS 353
DB 300 VERKYDLSHLKEIASGAPLSKEI-----GEMVKKRFLNFRQGYGLTETTS 347
QY 354 CFTTTPPLSMAGYIGLPAPCEYKLV-PVDGKLEG-----RFHGHVNSGYWRAPQON 405
DB 348 VLTTPKGDAPGSTGKIVPRHAYKVVDPTTGKTLGNREPELTFKGMIMKGYNNBEAT 407
QY 406 AQAFDEGYCSDAIKLADPADPQKGLMFDGRIABDFKLSGCVFVSGVGLRTFAVLEGG 465
DB 408 KAIIDNGMLRSGDI-----AYYD-----NDGHFYIVDLKSLIKYKG- 445
QY 466 SYLDDVVAAPDEECGLVFPRLDORALSGLEASDAEVLASEPRVAMPDMLKRLN 525
DB 446 ---YQVAPAEIEGI-LLOHPYIVDA-GVTGIPDEAGELPAAGVVVQGT-----KYLN 493
QY 526 RE-----ATGNASRIW-----VGLDTPPSIDKGEVTDK 555
DB 494 EQIVQDFVSSQVSTAKMLRGVAFKFLDEIPKSGTGKIDRK 532

RESULT 14
US-09-396-154-20
; Sequence 20, Application US/09396154
; Patent No. 6602677
; GENERAL INFORMATION:
; APPLICANT: Wood, Keith V.
; APPLICANT: Hall, Mary P.
; TITLE OF INVENTION: Thermostable luciferases and methods of
; FILE REFERENCE: 341.012051
; CURRENT APPLICATION NUMBER: US/09/396.154
; CURRENT FILING DATE: 1999-09-15
; EARLIER APPLICATION NUMBER: US 09/156.946
; EARLIER FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: PCT/US98/19494
; EARLIER FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: US 60/059.379
; EARLIER FILING DATE: 1997-09-19
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 20
; LENGTH: 544
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Mutant luciferase
; NAME/KEY: UNSURE
```

; LOCATION: (354)...(355)
; OTHER INFORMATION: Unknown amino acids
US-09-396-154-20

Query Match 8.2%; Score 251.5; DB 4; Length 544;
Best Local Similarity 22.8%; Pred. No. 1.2e-17;
Matches 131; Conservative 90; Mismatches 225; Indels 129; Gaps 26;

```
QY 30 CVAARANGEMRISYAEMFHVRAIAQSLPFGLSAEPRLLVSGNDLEHLQAFGAMY 89
DB 38 CIALTNHNR-ENVLYEEFLKSCRLAESFKKYGKLNQNTIAVCSENGLOFLPVIASLY 96
QY 90 AGIPYCPVSPAYSLSDILAKLRIHVGLQPLVFADAAAPFORA-----ITFLP 140
DB 97 LGIIAAPVSDKY--IERELI---HSLGIKPRILIFCSKNT-FQKVLNVSCKLYVETIIL 150
QY 141 DVVPAIFTEGELAGRTVSPDLSLEPGIIEAD---NAPATGPRTIAKFLFTSGSTKL 196
DB 151 LDI-----NEDLGVCYCL--NNFTSQNSDINLDVKKFKPYSFNRDQVALVMESSGTTG 203
QY 197 PKAVPTTQMLCANQOMLQTPFVFG---EEPRVLDVLPVMTTFGSHNIG-----I 246
DB 204 PKGVMLTHNIVA--RFSIAKQDPTFGNAINPTTALITVLPFHGFGMMTTLGFTGFRV 261
QY 247 VLYNGSTYYLDGKPTAOGFAET--LRNLSEISPTAYLTVPKGMELVGALEERDSTLR 304
DB 262 VLMH-----TFEEKLFLQSLQDYKVESTLVP-----TLMAFLAKSALVERX 303
QY 305 FPARMKLFPPAAAGLSQGIWDRIDRVABQHCGRIR-----MMAGLMTEPTAPSCT 357
DB 304 DUSHLKEIASGAPLSKEI-----GEMVKKRFLNFRQGYGLTETTSVLT 351
QY 358 TGPLSMAGYIGLPAPCEYKLV-PVDGKLEG-----RFHGHVNSGYWRAPQONQAF 409
DB 352 PKXXVAPGSTGKIVPRHAYKVVDPTTGKTLGNREPELTFKGMIMKGYNNBEATKAI 411
QY 410 DEEGYCSGDAIKLADPADPQKGLMFDGRIABDFKLSGCVFVSGVGLRTFAVLEGGSYL 469
DB 412 DKDGMRLSGDI-----AYYD-----NDGHFYIVDLKSLIKYKG----- 445
QY 470 DVVVAAPDEECGLVFPRLDORALSGLEASDAEVLASEPRVAMPDMLKRLREAT 529
DB 446 ---YQVAPAEIEGI-LLOHPYIVDA-GVTGIPDEAGELPAAGVVVQGT-----KYLNEQIV 497
QY 530 GN-----ASRIW-----VGLDTPPSIDKGEVTDK 555
DB 498 QNFVSSQVSTAKMLRGVAFKFLDEIPKSGTGKIDRK 532

RESULT 15
US-09-602-628-2
; Sequence 2, Application US/09602628
; Patent No. 6495355
; GENERAL INFORMATION:
; APPLICANT: Bames, Brian
; APPLICANT: Contag, Christopher
; TITLE OF INVENTION: Red-Shifted Luciferase
; FILE REFERENCE: SUN-127
; CURRENT APPLICATION NUMBER: US/09/602.628
; CURRENT FILING DATE: 2000-06-21
; PRIOR APPLICATION NUMBER: 60/140.598
; PRIOR FILING DATE: 1999-06-22
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 550
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Red shifted luciferase mutation
US-09-602-628-2

Query Match 8.1%; Score 249.5; DB 4; Length 550;
```

Best Local Similarity 24.9%; Pred. No. 2e-17;
Matches 104; Conservative 67; Mismatches 161; Indels 85; Gaps 17;

```
QY 43 ISTAEMFHNVRALIAQSLLPYGLSARPLLYSGNDLHLQAFGAMYAGIPYCPVSPAYS 102
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 51 ITVAEYFEMSGVRLAEAMKRYGLNTNHRIVWCSENSIQFMPVLGALFIVAVAPANDIY- 109
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 103 LLSODLAKLHHIVGLQPGLVFADAPFQRAIETIL--PDDVPAITTRGELAGRTVSF 160
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 110 -----NERELINSMGISQPTVFVS---KKGLQKILNVQKLPIT-----QKIIIM 152
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 161 DSLLE-----QPGGIEADNAFAATGPD-TIAKFLTSGSTKLPKAVPTTQ 204
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 153 DSKTDYQGFQSGMYTFVTSHLPFGFNEYDFVPESEFDKDTITALIMNSSGSTGLPKGVALPH 212
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 205 RMLCANQOMLLQTFPVFGEE--PVLVDMLPNNHTFGGSHNIG-----IVLYNGT 253
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 213 RTACVRFSSHARD--PIFGNOIIPDTAILSVVPFHGFGMFTTLGYLICGPRVVLMYR--- 267
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 254 YVLDGKPTAGFAET--LNLSEISPTAVLTVPKGMELVGALERDSTLREF-FARMK 310
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 268 -----FEEELPLRSLODYKTQSALVP---TLSPFAKSTLIDKYDLSNLH 310
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 311 LFFFAAGLSQGIWDRIDRVAEOHCHGERIRMMAGLGMTETAPSCFTTGPLSMAGYIGLP 370
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 311 EIASGGAPLSKEI---SEAVAKRFLHPGIRQ--GYGLTETTSALILITPEGDDKPGAYGV 365
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 371 APGCEVKLVVD-GKLEG-----RFGPHVMSGVWRAPQNAQAFDEEGYCSGD 419
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 366 VPFEFAKVVDLDGKTLGVNQRGELCVRGPMIMSGYNNPEATNALIDKGMLHSGD 422
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: February 7, 2005, 23:50:55
Job time : 46 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2005, 21:35:24 ; Search time 113 Seconds
(without alignments)
1869.837 Million cell updates/sec

Title: US-09-750-986D-30
Perfect score: 3082
Sequence: 1 MRLSLALLPRLPGRILRLERH.....KVDAIYRGDSQMLRDEATL 569

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729239 residues
Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_23Sep04:.*
1: geneseqp1980s:.*
2: geneseqp1980s:.*
3: geneseqp2000s:.*
4: geneseqp2001s:.*
5: geneseqp2002s:.*
6: geneseqp2003as:.*
7: geneseqp2003bs:.*
8: geneseqp2004s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1652.5	53.6	633	6	ADA35232 Actinobac
2	775	25.1	423	3	AAy87970 Pseudomon
3	716.5	23.2	336	3	AAy87971 Pseudomon
4	698.5	22.7	161	3	AAy87972 Pseudomon
5	312.5	10.1	607	5	ABP65419 Bifidobac
6	300.5	9.8	648	5	AAU44298 Propionib
7	300.5	9.8	648	6	ABM40817 Propionib
8	300	9.7	457	7	ADB74303 Mycobacte
9	298	9.7	488	7	ADB74534 Mycobacte
10	283	9.2	700	5	ADL23382 Plant acy
11	283	9.2	700	7	ADL72359 Modified
12	283	9.2	700	8	ADN72469 Thale cre
13	282	9.1	700	7	ADL72422 Arabidops
14	279	9.1	700	7	ADL72260 Arabidops
15	278.5	9.0	695	5	ABP66163 Bifidobac
16	278	9.0	701	5	ADL23381 Plant acy
17	278	9.0	701	7	ADL72259 Arabidops
18	275	8.9	701	7	ADL72421 Arabidops
19	266	8.6	614	7	ADP66781 Bacterial
20	265	8.6	545	7	ADP66860 Mutant lu
21	265	8.6	6842	8	AD139289 S. hygro
22	264.5	8.6	546	2	AAy06859 Mutant lu
23	263.5	8.5	544	4	AAU01228 Beetle lu
24	263	8.5	634	8	ADM86620 Human VLC
25	263	8.5	634	8	ADM16997 Human VLC

26	263	8.5	663	8	ADM16999 Human VLC
27	263	8.5	724	6	ABP57717 Novel hum
28	262.5	8.5	544	4	AAU01229 Beetle lu
29	262.5	8.5	544	4	AAU01233 Beetle lu
30	262.5	8.5	544	5	ABP53346 Photuris
31	262.5	8.5	544	5	ABP53344 Photuris
32	262.5	8.5	544	8	ADM32121 Mutant lu
33	262.5	8.5	544	8	ADM32119 Mutant lu
34	262.5	8.5	4317	6	ABU39938 Protein e
35	260	8.4	628	4	ABU3215 Human met
36	257.5	8.4	544	4	AAU01232 Beetle lu
37	257.5	8.4	544	5	ABP53345 Photuris
38	257.5	8.4	544	8	ADM32120 Mutant lu
39	256	8.3	536	6	ABU15399 Protein e
40	256	8.3	536	8	ADK13781 E. coli i
41	255.5	8.3	544	4	AAU01226 Beetle lu
42	255.5	8.3	546	2	AAy06857 Mutant lu
43	254.5	8.3	544	4	AAU01227 Beetle lu
44	254.5	8.3	546	2	AAy06858 Mutant lu
45	253	8.2	662	4	AAU39437 Propionib

ALIGNMENTS

RESULT 1
ADA35232 standard; protein; 633 AA.
ID ADA35232
XX
AC ADA35232;
XX
DT 20-NOV-2003 (first entry)
XX
DE Acinetobacter baumannii; bacterial disease; antibacterial; vaccine;
XX
KW Acinetobacter baumannii; bacterial disease; antibacterial; vaccine;
KM plant biocontrol agent.
XX
OS Acinetobacter baumannii.
XX
PN US6562958-B1.
XX
PD 13-MAY-2003.
XX
PP 04-JUN-1999; 99US-00328352.
XX
PR 09-JUN-1998; 98US-0088701P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Breton G, Bush D;
XX
DR WPI; 2003-576092/54.
XX
N-PSDB; ADA31106.
XX
PT New Acinetobacter baumannii proteins and nucleic acids, useful as reagents
PT for diagnosing a bacterial disease, as components of antibacterial
PT vaccines, as targets for antibacterial drugs, or as biocontrol agents for
PT plants.
XX
PS Example; SEQ ID NO 6519; 328pp; English.
XX
CC The invention relates to isolated Acinetobacter baumannii; nucleic acids.
CC The A. baumannii nucleic acids and polypeptides are useful as reagents
CC for diagnosing a bacterial disease, as components of antibacterial
CC vaccines, as targets for antibacterial drugs, to detect the presence of
CC A. baumannii and other Acinetobacter species in a sample, in screening
CC compounds for the ability to interfere with the A. baumannii life cycle
CC or to inhibit A. baumannii infection, and as biocontrol agents for
CC plants. The present sequence represents the amino acid sequence of an A.
CC baumannii protein.
XX
SQ Sequence 633 AA:

Query Match	53.6%;	Score 1652.5;	DB 6;	Length 633;
Best Local Similarity	53.4%;	Pred. No. 1.9e-152;		
Matches 310;	Conservative 100;	Mismatches 168;	Indels 3;	Gaps 3;

QY 5 FALLPEPRILIREHMAKTRPEQCVARANGERRISYEMFNNVALIQSOLLPYGI 64

Db 45 EOLKPEPKLIDRLIHFQOTKPDHILPAARNGOEVNVLKSYAEVLORAWHIALPARNI 104

QY 65 SAERPLILVSGNDLEHLQIAFGAMVAGIPYCVSPAYSILSODLAKIRI VGLIOPGLVF 124

Db 105 SQERPLIVLSGNDLEHLITLSMAAMLAGVFSALSPAYSILSODFGKLKXIVFVLTGMY 164

QY 125 AADAPEFORALETIIPDVPALETFGELAGRRTVSPDSILBOPGIEADNAFAATPDTI 184

Db 165 ASDQAFKAKIAQACTIPDIEVTVNKGIVGDOI.CTFSQSLDTPVFS-NVOBFOYTLDENOI 223

QY 185 AKPLFTSSTKLPRKVPPTTQMLCNOQMLLOTFFPGSEPLVNLVNMNHTFGSSHN 244

Db 224 AKPLFTSSTKLPRKVPPTTHMLCNOQMLLOTFFPEETPLVNLVNMNHTFGSSHN 283

QY 245 GIVLVNGTYYLDDOKPTAOGFAETLRNLSEISPTAYLVPKGMBEVLALERDSTLRER 304

Db 284 GIALYNGTIIYDDOKPVAKGTDEIIRNLKESIPTYLVNPKGMBEVLTALERDSTLRER 343

QY 305 FFAARKLFFPAAAGLSOGIWDRLDVAEOHCGERIRMMAGLQMTETAPSCFTTGPLSMA 364

Db 344 FFAKRLIFFGAALSEGMNRLDRIAQOHCSEKIRIMGLQMTETAPSCAFTTGPVMA 403

QY 365 GYIGPAAGCEVKLVPVUDKLEGRFHGPHVMGMYR-ABEQNAQADDEGYCSDAITKL 423

Db 404 GFIGPAGCEIKLVPCGDKLECYVRGKVMGMYRMLKADQOSTIFFDDEGFYTGAVRL 463

QY 424 ADPAPOGGLMFEDGRIAEDFKLSGVFVSGPLRPAVLVEGGSYLVVVAAPDRECLGI 483

Db 464 VDVPDPTGLMWDGRIAEDFKLTNTGTPVAVGLTRKLVLIQMLLIDVCITGSNLAIGF 523

QY 484 LVFPLLDORALSGIG-KEASDAEVLASEPVAWPAWMLKRLNREATGNASRIMVGLLD 542

Db 524 LIFPLDACAQYAGKLGEYSAAETILQHPKVQOWFROFLTTFNKDATGSSNTVSMLYMT 583

QY 543 TPBSIDKEBYVDKGINORAVLIQMSAKKDALYRGEDGML 583

Db 584 EPPQIDAEVTDKNIQSSIRKRAALIDELIYOKOTDPL 624

RESULT 2
AAV87970
ID AAV87970 standard; protein; 423 AA.

DT 21-SEP-2000 (first entry)

DE *Pseudomonas* sp. HR199 fcs-omega-Km protein.

KM Eugenol; ferulic acid; coniferyl alcohol; coniferyl aldehyde; vanillin;
KM vanillic acid; flavouring; ferulic acid-CoA-synthetase; fcs;
KM fcs-omega-Km.

OS Pseudomonas sp.

PN DE19850242-A1.

PD 04-MAY-2000.

PF 31-OCT-1998; 98DE-01050242.

PR 31-OCT-1998; 98DE-01050242.

PA (HAAR) HAARMANN & REIMER GMBH.

PI Rabenhorst J, Steinbuechel A, Priefert H, Overhage J;

XX WPI: 2000-340642/30.
DR
DR N-PSDB; AAA39720.
XX
XX Transformed organisms, used for production of vanillin and other
PT methoxyphenols, have altered catabolism of eugenol and ferulic acid.
XX
XX Disclosure: Fig 2g; 80pp; German.

This invention describes novel transformed and/or mutated uni- or multi-cellular organisms (A) in which enzymes (I) involved in catabolism of eugenol (B) and/or ferulic acid (FA) are inactivated in such a way that the organism accumulates the intermediates coniferyl alcohol (CAI), coniferyl aldehyde (CA2), FA, vanillin (V) and/or vanillinic acid (VA). INDEPENDENT CLAIMS are also included for the following: (a) gene structures in which sequences that encode one or more of the enzymes CAI-dehydrogenase, CA2-dehydrogenase, FA-CoA-synthetase, enoyl-CoA hydratase-aldoalase, beta-ketochiolase, V-dehydrogenase and VA-demethylase have been altered and/or inactivated; (b) vector containing at least one gene structure of (a); (c) biotechnical production of organic compounds (particularly alcohols, aldehydes or organic acids) using (A); and (d) production method of (A). (A) are used to prepare organic compounds, preferably alcohols, aldehydes and organic acids, especially CAI, CA2, FA, V and/or VA, most particularly V, a well-known natural flavouring. The inexpensive, starting material eugenol can be converted to V in a single step. This sequence represents the *Pseudomonas* sp. HR199 ferulic acid-CoA-synthetase, fcs-omega-Km which is described in the method of the invention

Sequence 423 AA;

Query Match	25.1%	Score 775;	DB 3:	Length 423;
Best Local Similarity	37.4%	Pred. No. 1.2e-66;		
Matches 231; Conservative	43;	Mismatches 121;	Indels 222;	Gaps 25;

Qy	1	MSLEKLLPFGSIIERLEHMAKTRPEOTCYAARANEMRISYAEHFNVRVIAQSL	60
Qy	2	MSLEKLLPFGSIIERLEHMAKTRPEOTCYAARANEMRISYAEHFNVRVIAQSL	60
Db	1	MSLEKLLPFGSIIERLEHMAKTRPEOTCYAARANEMRISYAEHFNVRVIAQSL	60
Qy	61	PGTSLERPLLVSQNDLHDLQAGMYAGIPCPSPASLSLSDATKLRHVLGLOP	120
Db	61	PGTSLERPLLVSQNDLHDLQAGMYAGIPCPSPASLSLSDATKLRHVLGLOP	120
Qy	121	GIWFAADAAFPORAIETLLPDDVPAITFRGELAGRTVSPDSLQPGIEADNAFAATG	180
Db	121	GIWFAADAAFPQ-----MEEDG-----	138

QY 181 PDITAKFLETSSTKLPKAVPTTQMLCANQOMLQTFVFGESEPVLDMLPMNHTFGG 240

[illegible]

1. The first step in the process of creating a new product is to identify a market need. This involves conducting market research to understand the preferences and behaviors of potential customers. Once a need is identified, the next step is to develop a concept that addresses this need. This concept should be unique and offer a clear value proposition to the target market.

[illegible]

THE UNIVERSITY OF CHICAGO PRESS

$\frac{1}{2}$
 $\frac{1}{3}$
 $\frac{1}{4}$
 $\frac{1}{5}$
 $\frac{1}{6}$
 $\frac{1}{7}$
 $\frac{1}{8}$
 $\frac{1}{9}$
 $\frac{1}{10}$
 $\frac{1}{11}$
 $\frac{1}{12}$
 $\frac{1}{13}$
 $\frac{1}{14}$
 $\frac{1}{15}$
 $\frac{1}{16}$
 $\frac{1}{17}$
 $\frac{1}{18}$
 $\frac{1}{19}$
 $\frac{1}{20}$
 $\frac{1}{21}$
 $\frac{1}{22}$
 $\frac{1}{23}$
 $\frac{1}{24}$
 $\frac{1}{25}$
 $\frac{1}{26}$
 $\frac{1}{27}$
 $\frac{1}{28}$
 $\frac{1}{29}$
 $\frac{1}{30}$
 $\frac{1}{31}$
 $\frac{1}{32}$
 $\frac{1}{33}$
 $\frac{1}{34}$
 $\frac{1}{35}$
 $\frac{1}{36}$
 $\frac{1}{37}$
 $\frac{1}{38}$
 $\frac{1}{39}$
 $\frac{1}{40}$
 $\frac{1}{41}$
 $\frac{1}{42}$
 $\frac{1}{43}$
 $\frac{1}{44}$
 $\frac{1}{45}$
 $\frac{1}{46}$
 $\frac{1}{47}$
 $\frac{1}{48}$
 $\frac{1}{49}$
 $\frac{1}{50}$
 $\frac{1}{51}$
 $\frac{1}{52}$
 $\frac{1}{53}$
 $\frac{1}{54}$
 $\frac{1}{55}$
 $\frac{1}{56}$
 $\frac{1}{57}$
 $\frac{1}{58}$
 $\frac{1}{59}$
 $\frac{1}{60}$
 $\frac{1}{61}$
 $\frac{1}{62}$
 $\frac{1}{63}$
 $\frac{1}{64}$
 $\frac{1}{65}$
 $\frac{1}{66}$
 $\frac{1}{67}$
 $\frac{1}{68}$
 $\frac{1}{69}$
 $\frac{1}{70}$
 $\frac{1}{71}$
 $\frac{1}{72}$
 $\frac{1}{73}$
 $\frac{1}{74}$
 $\frac{1}{75}$
 $\frac{1}{76}$
 $\frac{1}{77}$
 $\frac{1}{78}$
 $\frac{1}{79}$
 $\frac{1}{80}$
 $\frac{1}{81}$
 $\frac{1}{82}$
 $\frac{1}{83}$
 $\frac{1}{84}$
 $\frac{1}{85}$
 $\frac{1}{86}$
 $\frac{1}{87}$
 $\frac{1}{88}$
 $\frac{1}{89}$
 $\frac{1}{90}$
 $\frac{1}{91}$
 $\frac{1}{92}$
 $\frac{1}{93}$
 $\frac{1}{94}$
 $\frac{1}{95}$
 $\frac{1}{96}$
 $\frac{1}{97}$
 $\frac{1}{98}$
 $\frac{1}{99}$
 $\frac{1}{100}$

CONSTITUTIONAL COURT

\vdots

[illegible][illegible][illegible][illegible]

Db	336	LCGEWADRVLVYIGIAAPDSQRIAFYRLD-----EFFVLQWRSKVD	408
Qy	573	ALYRGEDQSMURDEATL	589
Db	407	ALYRGEDQSMURDEATL	423
RESULT 3			
ID	AAy87971		
XX	AAy87971	standard; protein; 336 AA.	
XX	AAy87971;		
XX	AC		
XX	DT	21-SEP-2000 (first entry)	
XX	DE		
XX	DE	Pseudomonas sp. HR199 fcs-omega-Gm protein.	
KM	KM	Eugenol; ferulic acid; coniferyl alcohol; coniferyl aldehyde; vanillin;	
KW	KW	vanillinic acid; flavouring; ferulic acid-CoA-synthetase; fcs;	
XX	XX	fcs-omega-Gm.	
OS	OS	Pseudomonas sp.	
XX	XX	DE19850242-A1.	
XX	XX	04-MAY-2000.	
XX	XX	31-OCT-1998; 98DE-01050242.	
XX	XX	31-OCT-1998; 98DE-01050242.	
XX	XX	(HAAR) HAARMANN & REIMER GMBH.	
PI	PI	Rabenhorst J, Steinduechel A, Priefert H, Overhage J;	
XX	XX	WPI; 2000-340642/30.	
DR	DR	N-PSDB; AAA39721.	
PT	PT	Transformed organisms, used for production of vanillin and other	
XX	XX	methoxyphenols, have altered catabolism of eugenol or ferulic acid.	
XX	XX	Disclousure; Fig 2h; 80pp; German.	
XX	XX		
CC	CC	This invention describes novel transformed and/or mutated uni- or multi-	
CC	CC	cellular organisms (A) in which enzymes (I) involved in catabolism of	
CC	CC	eugenol (E) and/or ferulic acid (FA) are inactivated in such a way that	
CC	CC	the organism accumulates the intermediates coniferyl alcohol (CA1),	
CC	CC	coniferyl aldehyde (CA2), FA, vanillin (V) and/or vanillinic acid (VA).	
CC	CC	INDEPENDENT CLAIMS are also included for the following: (a) gene	
CC	CC	structures in which sequences that encode one or more of the enzymes CA1-	
CC	CC	aldolase, CA2-dehydrogenase, FA-CoA-synthetase, enoyl-CoA hydratase-	
CC	CC	aldolase, beta-ketothiolase, V-dehydrogenase and VA-demethylase have been	
CC	CC	altered and/or inactivated; (b) vector containing at least one gene	
CC	CC	structure of (a); (c) biotechnical production of organic compounds	
CC	CC	(particularly alcohols, aldehydes or organic acids) using (A); and (d)	
CC	CC	production method of (A). (A) are used to prepare organic compounds,	
CC	CC	preferably alcohols, aldehydes and organic acids, especially CA1, CA2,	
CC	CC	FA, V and/or VA, most particularly V, a well-known natural flavouring.	
CC	CC	The inexpensive starting material eugenol can be converted to V in a	
CC	CC	single step. This sequence represents the Pseudomonas sp. HR199 ferulic	
CC	CC	acid-CoA-synthetase, fcs-omega-Gm which is described in the method of the	
CC	CC	invention	
XX	XX		
XX	XX	Sequence 336 AA;	

```

Qy      61 PYGSAEPRLLVSGNDDEHLOLAGAWAGI PYCPVSPAYSLSDOLAKRHIYELLPO 120
Db      61 PPGLSAERBLTVSGNDDEHLOLAGAWAGI PYCPVSPAYSLSDOLAKRHIYELLPO 120
Qy      121 GLVFAADAAAPFORAIEITLLPDVPAIPTF--GELAGRRTVSPFDSLDECGGIEADNAFMA 178
Db      121 GLVFAADAAAPFQMRRSS--NDVTQGSRPKTKLG-----SSMGIIRT-----CR 163
Qy      179 TGPDITAKFLTFSSGSKLPKAVPTTQRMLCANQOMLOTFPPYFGGEPRPLVDMLPNNHTF 238
Db      164 LGPDGV-----KSMRAALDLFGRE-----F 183
Qy      239 GGSNIGIVLYNGGYTLDDGKPTAQGAETLRNLSEISPTAYLTVPKMGEIYVGLAERD 298
Db      184 G-----DVATY--SQHPDSDDYLGMLRSTFTLAAAF-----DQEAIVGAL----- 223
Qy      299 STLREFFRARKMLFFFAAAGLSQGIWDRIDLRVAEQHCERIRMMAGLCTETAPSTCTTT 358
Db      224 ----- 223
Qy      359 GPLSMAGYIGLPAPGCEVTLVPVDGKLEGRFPHGPMGSGYRAPAEONAPDEEGYCSG 418
Db      224 -----AAVY-LP-----RFGQPR-----SEIT----- 239
Qy      419 DAIKLADPADPOKGLMFDGRINAEFRKLSGVSFVSGLPFTRAVLBEGSYVLDVVAAPPR 478
Db      240 -----IYDLAVSGEHR--ROGITATALINLKHEANMLGAVVIYVQADYGD 283
Qy      479 ECLGLVFPRLDRCRALGLGKEAGDAEVLASEPRAWADMKRLNREATGNASRIMVY 538
Db      284 PAVAL--YTKL-----GIRBE-----V 298
Qy      539 GLDTPPSIDKGEVTDKGSINQRAVLQWRSAYDALYRGEDOSMLRDEATL 589
Db      299 MHFDIDPS-----TATVLQWSACVDALYRGEDOSMLRDEATL 336

```

[illegible]

CC cellular organisms (A) in which enzymes (I) involved in catabolism of
CC eugenol (B) and/or ferulic acid (FA) are inactivated in such a way that
CC the organism accumulates the intermediates confertyl alcohol (CAI),
CC confertyl aldehyde (CA2), FA, vanillin (V) and/or vanillic acid (VA).
CC INDEPENDENT CLAIMS are also included for the following: (a) gene
CC structures in which sequences that encode one or more of the enzymes CAI-
CC dehydrogenase, CA2-dehydrogenase, FA-CoA-synthetase, enoyl-CoA hydratase-
CC aldolase, beta-ketothiolase, V-dehydrogenase and VA-demethylase have been
CC altered and/or inactivated; (b) vector containing at least one gene
CC structure of (a); (c) biotechnical production of organic compounds
CC (particularly alcohols, aldehydes or organic acids) using (A); and (d)
CC production method of (A). (A) are used to prepare organic compounds,
CC preferably alcohols, aldehydes and organic acids, especially CAI, CA2,
CC FA, V and/or VA, most particularly V, a well-known natural flavouring.
CC The inexpensive starting material eugenol can be converted to V in a
CC single step. This sequence represents the Pseudomonas sp. HR199 ferulic
CC acid-CoA-synthetase, fcs-delta which is described in the method of the
CC invention

XX Sequence 161 AA;

Query Match 22.7%; Score 698.5; DB 3; Length 161;

Best Local Similarity 92.1%; Pred. No. 9e-60; Mismatches 8; Indels 1; Gaps 1;

Matches 139; Conservative 3; Mismatches 8; Indels 1; Gaps 1;

QY 1 MRSLEALPPGRIERLEHMAKTRPEQTCVAAARAANGEMRRISYAEHFNVAIAQSLL 60
DB 1 MRSLEALPPGRIERLEHMAKTRPEQTCVAAARAANGEMRRISYAEHFNVAIAQSLL 60

QY 61 PYGLSERPLLYSGNDLEHQLAFGAMVAGIYCVSPAYSLLSODLAKRIHVLGP 120
DB 61 PYGLSERPLLYSGNDLEHQLAFGAMVAGIYCVSPAYSLLSODLAKRIHVLGP 120

QY 121 GLVFAADAAFPORAFETILPDVPAIFTRGE 151
DB 121 GLVFAADAAFPORAFETILPDVPAIFTRGE 150

RESULT 5
ABP65419
ID ABP65419 standard; protein; 607 AA.

XX ABP65419;
XX
XX 19-NOV-2002 (first entry)

DE Bifidobacterium longum NCC2705 ORF amino acid sequence SEQ ID NO.163.

XX Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;
XX anti-diarrhetic; antibacterial; inhibitor of Salmonella; detection;
XX identification; lactic acid bacterium; diarrhoea; pathogenic bacteria;
XX rotavirus; food composition; pharmaceutical composition.

OS Bifidobacterium longum.

XX EPI227152-A1.

XX 31-JUL-2002.

XX 30-JAN-2001; 2001EP-00102050.

XX 30-JAN-2001; 2001EP-00102050.

XX (NEST) SOC PROD NESTLE SA.

XX WPI; 2002-668397/72.

XX Novel polynucleotide comprising Bifidobacterium genome sequence useful as
XX a probe or primer for detecting and/or identifying Bifidobacterium longum
XX in a biological sample.

XX Claim 3; SEQ ID NO 163; 80pp; English.

CC The present invention describes a polynucleotide (I) comprising a
CC sequence of a Bifidobacterium genome selected from the nucleotide
CC sequence given in AB081842 and AB081843, or a sequence exhibiting at
CC least 90% identity or which hybridises with the sequences given in
CC AB081842 and AB081843. Also described is a polynucleotide (II) encoding a
CC fusion protein, comprising a sequence selected from 1097 sequences given
CC in ABP65258 to ABP65354 ligated in frame to a polynucleotide encoding a
CC heterologous polypeptide. (I) has anti-diarrhetic and antibacterial
CC activities, and can be used as an inhibitor of Salmonella. (I) (which is
CC a probe) is useful for the detection and/or identification of
CC Bifidobacterium longum in a biological sample. A carrier containing the
CC lactic acid bacterium Bifidobacterium longum NCC2705 (NCIM I-2618) can be
CC used for preventing and/or treating diarrhoea brought about by pathogenic
CC bacteria and/or rotavirus. The carrier is a food composition selected
CC from milk, yogurt, curd, cheese, fermented milks, milk based fermented
CC products, ice-creams, fermented cereal based products, milk based
CC powders, infant formula, pet food or a pharmaceutical composition
CC selected from tablets, liquid bacterial suspensions, dried oral
CC supplement, wet oral supplement, dry tube feeding or wet tube feeding.
CC (1) is useful in DNA arrays or chips to carry out analysis of the
CC expression of the Bifidobacterium gene. AB081844 to AB081850 represent
CC Bifidobacterium related nucleotide sequences given in the Sequence
CC listing from the present invention but not mentioned further within the
CC specification. N.B. The sequence data for this patent is not represented
CC in the printed specification but is based on sequence information
CC supplied by the European Patent Office

XX Sequence 607 AA;

Query Match 10.1%; Score 312.5; DB 5; Length 607;

Best Local Similarity 23.5%; Pred. No. 5.2e-21; Mismatches 245; Indels 129; Gaps 21;

Matches 142; Conservative 87; Mismatches 245; Indels 129; Gaps 21;

QY 14 IERLEHMAKTRPEQTCVAAARAANGEMRRISYAEHFNVAIAQSLLPYGLSERPLLY 73
DB 24 IERLEHMAKTRPEQTCVAAARAANGEMRRISYAEHFNVAIAQSLLPYGLSERPLLY 80

QY 74 SGNDEHQLAFGAMVAGIYCVSPAYSLLSODLAKRIHVLGPGLVFAAD 127
DB 81 SGNDEHQLAFGAMVAGIYCVSPAYSLLSODLAKRIHVLGPGLVFAAD 129

QY 128 AAPQR-----ALETLPDDVPAIFTRGEIAGR-TVSRDLSLEPGIEANNAFA 177
DB 130 RERFRLDSVKDCHCPALKQILMEGNAL--GALBGLGVTSDELEHRAVATRTD--- 183

QY 178 ATGPDTIAKFLFTSGSTKCPKAVPTTQ-----MLCANQOMLQTPVGESEPVLDWL 232
DB 184 -----LATIVTSGSTGNPKGAEULTKNFVSTTISAQAL-----HEVLDDHRLRLFL 233

QY 233 PMNHTEGSGHNIGVLYNGTY-YLDDGKPTAGFAETLRNLSISPTAYLTVPKGMBEL 291
DB 234 PLAHCFARFIQVASIASDGGVGLPDTK-----TLPLDLSFPFTYLVGPVFEKV 286

QY 292 VGALERDSTLRERFPARMKLFPPAAGLSOGIWDRLDRVAEQH----- 334
DB 287 YNAASHKAG-----AGWKGRLFEVKAABARVWSRKEQGEQTFALIERAKETLVYR 340

QY 335 -----CGER-----IRWAGLMTETAPSCFTTGPPLSMAGY 366
DB 341 TVRGALGPRIKIVKACGAPLSDLAHFNVGIGLPMIOGIMETETAPPAATRVTDNVIGT 400

QY 367 IGLPAPCEVKLVVDGKLEGRFHPHVNWSGYWRABEONAQAFDEGGYCSGDAIKLADP 426
DB 401 VGOPAPGSSIR-ISDEGELQVK--GNVVRGYHNLPEKTAFAFTADGWLRTGDIAEI--- 454

QY 427 ADPOKGLMPDGRILAEFKLSGVFVSVGLRTRAVLEGSSYLVUDVVAAPDEBGLLYF 486
DB 455 -DDEGHIVTGIKIDITTAGKVNPSIPLEBE--TAKCPVHCVVADQRPFTALV- 510

QY 487 PRLDGRALS-----GLGKEASDAEVLASEPRAWAFADWLKRLNBEATGNASRIMWGL 540
DB 511 --TLDESLALWLPAGHLSSTETPVDRLATNAAVREIQGYVDKAN-ATYSRAESVKKFAY 567

QY 541 LDT 543
DB 568 LDT 570

RESULT 6
AAU44298
AAU44298 standard; protein; 648 AA.

XX AAU44298;

XX 27-FEB-2002 (first entry)

XX Propionibacterium acnes immunogenic protein #5194.

XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
XX uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
XX inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
XX dermatological; osteopathic; neutroprotectant.

XX Propionibacterium acnes.

XX MO200181581-A2.

XX 01-NOV-2001.

XX 20-APR-2001; 2001MO-US012865.

XX 21-APR-2000; 2000US-0199047P.

XX 02-JUN-2000; 2000US-0208841P.

XX 07-JUL-2000; 2000US-0216747P.

XX (CORI-) CORIXA CORP.

XX Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhactia A;

XX L' Maisonneuve J, Zhang Y, Jen S, Carter D;

XX MPI; 2001-616774/71.

XX N-PSDB; AASS9522.

XX Example 1; SEQ ID NO 5493; 106ppp; English.

XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
XX polypeptides. The proteins and their associated DNA sequences are used in
XX the treatment, prevention and diagnosis of medical conditions caused by
XX P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
XX pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
XX P. acnes is also involved in infections of bone, joints and the central
XX nervous system, however it is particularly involved in the inflammatory
XX lesions associated with acne vulgaris. A method for detecting the
XX presence or absence of P. acnes in a patient comprises contacting a
XX sample with a binding agent that binds to the proteins of the invention
XX and determining the amount of bound protein in the sample. The
XX polypeptides may be used as antigens in the production of antibodies
XX specific for P. acnes proteins. These antibodies can be used to
XX downregulate expression and activity of P. acnes polypeptides and
XX therefore treat P. acnes infections. The antibodies may also be used as
XX diagnostic agents for determining P. acnes presence, for example, by
XX enzyme linked immunosorbent assay (ELISA). Note: The sequence data for
XX this patent did not form part of the printed specification, but was
XX obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences

XX Sequence 648 AA;

Query Match 9.8%; Score 300.5; DB 4; Length 648;
Best Local Similarity 23.0%; Pred. No. 8.6e-20;
Matches 151; Conservative 103; Mismatches 262; Indels 141; Gaps 28;

QY 9 PFGFR--ILRLBHWAKT-----RPEQTCVAAA-ANGEMRISYAEHFNVAIAQS 58
DB 27 PVGGENFIESHLAMRPAATVANGFRP-----ATRVAGGQWIIIRTAELGRRRAGLA 81
QY 59 LLPYGLSAERPL-----LIVGNDLEHLQLAFGAMYAGIPYCVSPAYSLLSODLAKR 112
DB 82 FVTPLGTTEDGLRGDRISLFGAGNCEPWIADLAGMTIGV--IPV--PIYPTSTDP--QIV 136
QY 113 HIYGLDQPLGVFPADAAFPQRALETILPDDVPAIFP-----RGLAGRRVSPDSL 163
DB 137 HIYTDAGVRVITTAGPRELDRIIEA--RDQMPGLEVILINPADQVGDHGLTVLSLEQ- 193
QY 164 LEOPGIE-----ADNAFAATGPDTIAKFLFTSGSTKLPRVAVPTQMLCANQOMLQTF 218
DB 194 VRQAGVSEELQYVBERMGSCPDVVAALITYSGTGERPGWIMSHQAALAE----LQAL 249
QY 219 PVFGEPPV--LVDMVPMNHTFGSGSHNIGIVLYNGTYYLDDGKPTAQGAETLRNLSI 276
DB 250 DAFEDVTVPADHSLFPLSHALEWGSMAVIRHGLCLNTFVNPRTIS-----AMLAEV 302
QY 277 SPTAVLTPKGMELVQALERDSTLERFF--ARMKLFPPA----- 315
DB 303 RPTLFVSVPRLYEQVM-----SVAREKVS DSPAKLIKFEWSIRIGEMWQAOEGRRPS 356
QY 316 -AAGLSQGIWDRP-----DRVAEQHCGERIRMMAGLGMTET 350
DB 357 VSLRAHGVADRVLVLAIRDAIGPKTVLAAGAPLKEVEFFPAACGLVCCQYGLTEA 416
QY 351 APSCTFTTGPLSMAGYIGLAPGCEVKLVVDGKLEGRFHPHVMGWRAPQONAFD 410
DB 417 SPLVSFNSPGYKFGYTAGRPVGSQMTTTE--DGEI--LYNGPMVMKGYKAPATAAI- 472
QY 411 EBGYCSGDAIKLADPADPQKGLMFDGRIADPKLSGVSFVSGPLTRAVLEGSVLVD 470
DB 473 EDGMLHRTDIGHI-----DEGDFLVITDRKLDIIVTGNKISQPLE-----NSLMKD 521
QY 471 -----VVVAAPDRRCGLLVFPRLDORAL-----SGLGEASDAEVLASEPVRWFA 518
DB 522 PLFEHVALGDNRPCLTLLVKSPLPOVEBLAEKHLHTSMGPMREBELAE-IRKRV 580
QY 519 DWLKRINREATGNASRIWVGLDTPPSIDKGEVTDKGSINQRAVLQWRSKVDALY 575
DB 581 EITEKLPHQEQIRDLRVLMDEF-----TTDNGLLTPTLKVRREVEKRFTEIYEMW 632

RESULT 7
ABM40817
ID ABM40817 standard; protein; 648 AA.
XX
AC ABM40817;
XX
XX 20-OCT-2003 (first entry)
XX
DE Propionibacterium acnes predicted ORF-encoded polypeptide #5493.
XX
XX Acne vulgaris; antiseborrheic; dermatological; antibacterial;
XX immunostimulant; immune response; vaccine.
XX
XX Propionibacterium acnes.
XX
XX MO2003033515-A1.
XX
XX 24-APR-2003.
XX
XX 11-OCT-2002; 2002MO-US032727.
XX
XX 15-OCT-2001; 2001US-00978825.
XX
XX (CORI-) CORIXA CORP.
XX
XX Skeiky YAM, Persing DH, Bhactia A, Maisonneuve JL;
XX Mitcham JL, Wang S, Jen S, Lodes WJ, Benson DR, Jones R, Carter D;
XX Zhang Y, Wang S, Jen S, Lodes WJ, Benson DR, Jones R, Carter D;
XX Barth B, Valliave-Douglas J;

XX WPI; 2003-381789/36.
DR N-PSDB; ACF64451.
XX
XX New Propionibacterium acnes polypeptides and polynucleotides encoding the
PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,
PT or for stimulating an immune response specific for a P. acnes protein.
XX
XX Example 1; SEQ ID NO 5493; 1481bp; English.
XX
CC The invention relates to an isolated polynucleotide (ACF64435-ACF64733)
CC encoding a Propionibacterium acnes protein. The invention also relates to
CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to
CC immunogenic fragments of P. acnes polypeptides. The invention
CC additionally encompasses expression vectors and host cells comprising a
CC polynucleotide of the invention; antibodies against polypeptides of the
CC invention; fusion proteins comprising a polypeptide of the invention; a
CC method for stimulating an immune response specific for a P. acnes
CC polypeptide and an isolated T cell population comprising T cells prepared
CC via this method; a vaccine composition (comprising P. acnes polypeptides,
CC polynucleotides, antibodies, fusion proteins, T cell populations, or
CC antigen-presenting cells that express the polypeptide); a method and kit
CC for detecting or determining the presence or absence of P. acnes in a
CC patient; and a method for inhibiting the development of P. acnes in a
CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion
CC proteins, T cell populations or antigen-presenting cells that express the
CC polypeptides are useful for diagnosing, preventing or treating acne
CC vulgaris, or for stimulating an immune response specific for a P. acnes
CC protein. The polynucleotides can also be used as probes or primers for
CC nucleic acid hybridisation. The vaccine composition is useful for the
CC stimulation of an immune response against P. acnes, or for treating acne,
CC and the kit is useful for performing a diagnostic assay. The present
CC sequence represents a polypeptide predicted to be encoded by an ORF (open
CC reading frame) contained within the P. acnes polynucleotides of the
CC invention. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIP0 at http://wipo.int/pub/published_pct_sequences
XX
SQ Sequence 648 AA;

Query Match 9.8%; Score 300.5; DB 6; Length 648;
Best Local Similarity 23.0%; Pred. No. 8.6e-20;
Matches 151; Conservative 103; Mismatches 262; Indels 141; Gaps 28;

QY 9 PEPGR--ILERLEHMAKT-----RPEQCVARA-ANGEMRISYAEWFHNVRAIAGS 58
DB 27 PVGEMFISHLAHMFRAVYANNGFRP-----ATRRQGQWILIRYAEIGRRVAGIARA 81
QY 59 LRPYGISAEKPL-----LIVSGNDLEHQLAFGANYAGIPYCPVSPANSLSLQDLAKR 112
DB 82 FVTPGILTEDGLQGRDRISLFAGNCEPWEIADLAGMTIGV--IPV-PIYPTSTPD--QIV 136
QY 113 HTVGLLOPGVAFADAAPFORALETILPDVPAIF-----RGLLAGRTVSPDSI 163
DB 137 HAVTDAGVRVITTAGKELEDRILEA--RDMPELETVILINPADQYGDHDLTVLSLEQ- 193
QY 164 LEQPGGIE-----ADNAFATGPDITAKFLFTSGSTKLPRKAVPTTORMCANQOMLLQTF 218
DB 194 VQAGVSEBIQTVBERMQSGCDVAALITSGTGEPRGWMISRAALAE-----LQAL 249
QY 219 PVFGEPRPV--LVDMLPMMHTFGSGSHNIGIVLYNGTYYIDGKPPAQAQFAETRLNLSI 276
DB 250 DAFFDVTAPDHSLSFPLSHALEMWSMAVIRHGCINTFVFNPKTIS-----AMLAIEV 302
QY 277 SPTAVLYTVKGMELGALERDSTLEKRF--AKMKLPFEA----- 315
DB 303 RPLFVSFKLYHQV-----SVAREKVS DSPAKKIFEWSSRRIGREMQAQBGRRS 356
QY 316 -AAGLSQGIWDRY-----DRVAEQHCEGRIRMAAGLMTET 350
DB 357 VSLRAHGVADRVLKAIKDAIGKPTVLAAGGAPLRKEVEBFPAACGLVCGGYGLTEA 416
QY 351 AFSCTTTGTGRLSMAGYIGLPAPGCEVLYPVVDGKLEGRFRHGHVMSGYMRAPRQNAQAPD 410

DB 417 SPLVSNFSPGGYKFFGAGRPVLSGQMTTTE-DGEI--LYRGPMVMGYMKAPATAAAI- 472
QY 411 EEGYCSGPAIKLADPADPOKGMFDPRIAEDEPKLSSGVFVSVPRTVAVLGGSGYVD 470
DB 473 EDOMLHTGIGHI-----DEDFGLVITDRKLDIIVLANGNISQPIE-----NSLMD 521
QY 471 -----VVVAAPRECEGLLVFPRLDCAI-----SGLGKASDAEVLASEPVWAPRA 518
DB 522 PLFEHAVLLGDNRPPCLTLVVKPSLPQVEELAEERLHTSMTGPEMLRSEELAE-IRRAVA 580
QY 519 DMLKLRNRAATGNASIMVWGLDTPSPIDKEVYDPKSGINQRAVLOMSAKYDALY 575
DB 581 ETTEKLPHEQJIRDLRVLDEF-----TTDNGLLTPTLVRRREVRKFTETVEEMY 632

RESULT 8
ADB74303
ID ADB74303 standard; protein; 457 AA.
XX
XX ADB74303;
AC
XX 04-DEC-2003 (first entry)
DT
XX
XX Mycobacterium leprae non-naturally occurring peptide #26.
DE
XX
XX Non-naturally occurring peptide; anion pump protein; tuberculosis;
KM hypersensitivity reaction; tuberculostatic.
XX
XX Mycobacterium leprae.
OS
XX
XX US6583266-B1.
PN
XX 24-JUN-2003.
PD
XX
XX 16-SEP-1994; 94US-00311731.
PF
XX
XX 19-AUG-1993; 93US-00109181.
PR
XX 22-OCT-1993; 93US-00142558.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
PA
XX Smith DR, Mao J;
PI
XX
XX WPI; 2003-656441/62.
DR N-PSDB; ADB74275.
DB
XX
PT New Mycobacterium tuberculosis anion pump peptide useful for as
tuberculosis vaccine and diagnosis of tuberculosis infection.
XX
PS Disclosure; SEQ ID NO 52; 26bp; English.
XX
CC The invention relates to a non-naturally occurring peptide of
CC Mycobacterium tuberculosis comprising an amino acid sequence
CC corresponding to an anion pump protein. The invention also relates to a
CC non-naturally occurring nucleic acid corresponding to a DNA sequence of
CC Mycobacterium tuberculosis or Mycobacterium leprae. The new peptide is
CC useful as a vaccine against Mycobacterium tuberculosis or Mycobacterium
CC leprae or for screening for new tuberculosis drugs. Purified proteins
CC derived from the sequences of the invention may elicit a specific immune
CC response. The peptide may also be used to detect hypersensitivity
CC reactions of individuals exposed to Mycobacterium tuberculosis or
CC Mycobacterium leprae. The proteins and peptides may be affixed to solid
CC supports to detect antibodies typical of hypersensitivity reactions, from
CC a patient's sera. This sequence represents Mycobacterium leprae non-
CC naturally occurring peptide of the invention. Note: The sequence data for
CC this patent did not form part of the printed specification but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 457 AA;

Query Match 9.7%; Score 300; DB 7; Length 457;

Best Local Similarity 26.2%; Pred. No. 5.7e-20;
Matches 143; Conservative 67; Mismatches 210; Indels 126; Gaps 27;
QY 44 SYAEFMFNVAIAQSLPFGISAEKRLIYSGNDLEHLQAFGAMVYGIPTCPVSPAYSL 103
DB 13 SVAEKVGARLVAVLATP---TASTVLATP-----GCLTIGVVPVPPADIGV 57
QY 104 LSQDLAKLRHIV-----GLLPGLVFAADAPFQRAIETILPDDVAIFRGLAGRT 157
DB 58 VER-----RMLTDSGAQAWLGPDP--SADSA-----PDGLPHI-----P 90
QY 158 VSFDSLLEPGGIEADNAFAATGPDITAKFLFTSGSTKLPRKAVPTTORMLCANQMLLOT 217
DB 91 VQDADA-----RSNMRYPEPSPDATAMVITYSGTGPCKGVLSSRAIYVDLALAQA 142
QY 218 FPFVGEPPVLVMDLPNNHTFG-----GSHNIGIVLYNGSTYYLDGCKPTAGCFATLR 271
DB 143 WQMTAAD--VLVHGLPFVHGLVGLIGLSLRI-----GNRFVHTGKPTPTAYQA-- 191
QY 272 NLSEISPTAYLTVPKGMEEVLGALERDSTLEREFARMKLEFFAAAGLSQIMDLRVA 331
DB 192 -CEBAGSLYFGVPTWSRLVA---DEAV-ARALRPARLLVSGSASLPVPVFDRLAHLT 245
QY 332 EOHGGERIRMMAGLGMTETAPSCFTTGPPLSMAGYIGLPAPGCEVKL-----VPYDGK 384
DB 246 GHRPIER-----YGSTESLITLSTLADGERRAGWGLPLAGVQTRLVDSESGFPYDGE 299
QY 385 LEBRF--HGFHWMSGYWRAPQNAQAFDEEGYCSGPAIKLADPADQKGLMFDGRJAED 442
DB 300 TVGRLOYRSPTMGGYLNPBEATAEAFDEBGMWRTGD--VAVVDSGGMHR--IYGRSVD 355
QY 443 FKLSGIVFVSGPLRTAVLEGSSVYLDV--VVAAPRECLGLVFPRLDORALSGLGE 501
DB 356 LKLGGRIGAGET--EMALLGHPDVREAVVGLPD--EDLQGRIVAFV-----VGAE 404
QY 502 ASDAEVLASBPVAMFADWMLKRLNREATGNASRIWVGLDTPPSIDKGEVTDKGSINOR 561
DB 405 ALDADELIN-----YVAQQLSIHKRPRE-----VRFDALPA--TRMRSRSSKSSCSR 450
QY 562 AVTLQMR 567
DB 451 A--DMR 454

RESULT 9
ID ADB74534 standard; protein; 488 AA.
AC ADB74534;
XX
DT 04-DEC-2003 (first entry)
DE Mycobacterium leprae non-naturally occurring peptide #234.
XX
KM Non-naturally occurring peptide; anion pump protein; tuberculosis;
XX hypersensitivity reaction; tuberculostatic.
OS Mycobacterium leprae.
PN US6583266-B1.
XX
PD 24-JUN-2003.
XX
PF 16-SEP-1994; 94US-00311731.
XX
PR 19-AUG-1993; 93US-00109181.
XX 22-OCT-1993; 93US-00142558.
PA (GENO-) GENOME THERAPEUTICS CORP.
XX Smith DR, Mao J;
XX PI
XX DR WPI; 2003-656441/62.

XX New Mycobacterium tuberculosis anion pump peptide useful for as
PT tuberculosis vaccine and diagnosis of tuberculosis infection.
PS Disclosure; SEQ ID NO 283; 26pp; English.
XX
CC The invention relates to a non-naturally occurring peptide of
CC Mycobacterium tuberculosis comprising an amino acid sequence
CC corresponding to an anion pump protein. The invention also relates to a
CC non-naturally occurring nucleic acid corresponding to a DNA sequence of
CC Mycobacterium tuberculosis or Mycobacterium leprae. The new peptide is
CC useful as a vaccine against Mycobacterium tuberculosis or Mycobacterium
CC leprae or for screening for new tuberculosis drugs. Purified proteins
CC derived from the sequences of the invention may elicit a specific immune
CC response. The peptide may also be used to detect hypersensitivity
CC reactions of individuals exposed to Mycobacterium tuberculosis or
CC Mycobacterium leprae. The proteins and peptides may be affixed to solid
CC supports to detect antibodies typical of hypersensitivity reactions, from
CC a patient's sera. This sequence represents Mycobacterium leprae non-
CC naturally occurring peptide of the invention. Note: The sequence data for
CC this patent did not form part of the printed specification but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 488 AA;
XX
Query Match 9.7%; Score 298; DB 7; Length 488;
Best Local Similarity 26.4%; Pred. No. 9.9e-20;
Matches 135; Conservative 61; Mismatches 196; Indels 120; Gaps 24;
QY 44 SYAEFMFNVAIAQSLPFGISAEKRLIYSGNDLEHLQAFGAMVYGIPTCPVSPAYSL 103
DB 50 SVAEKVGARLVAVLATP---TASTVLATP-----GCLTIGVVPVPPADIGV 94
QY 104 LSQDLAKLRHIV-----GLLPGLVFAADAPFQRAIETILPDDVAIFRGLAGRT 157
DB 95 VER-----RMLTDSGAQAWLGPDP--SADSA-----PDGLPHI-----P 127
QY 158 VSFDSLLEPGGIEADNAFAATGPDITAKFLFTSGSTKLPRKAVPTTORMLCANQMLLOT 217
DB 128 VQDADA-----RSNMRYPEPSPDATAMVITYSGTGPCKGVLSSRAIYVDLALAQA 179
QY 218 FPFVGEPPVLVMDLPNNHTFG-----GSHNIGIVLYNGSTYYLDGCKPTAGCFATLR 271
DB 180 WQMTAAD--VLVHGLPFVHGLVGLIGLSLRI-----GNRFVHTGKPTPTAYQA-- 228
QY 272 NLSEISPTAYLTVPKGMEEVLGALERDSTLEREFARMKLEFFAAAGLSQIMDLRVA 331
DB 229 -CEBAGSLYFGVPTWSRLVA---DEAV-ARALRPARLLVSGSASLPVPVFDRLAHLT 282
QY 332 EOHGGERIRMMAGLGMTETAPSCFTTGPPLSMAGYIGLPAPGCEVKL-----VPYDGK 384
DB 283 GHRPIER-----YGSTESLITLSTLADGERRAGWGLPLAGVQTRLVDSESGFPYDGE 336
QY 385 LEBRF--HGFHWMSGYWRAPQNAQAFDEEGYCSGPAIKLADPADQKGLMFDGRJAED 442
DB 337 TVGRLOYRSPTMGGYLNPBEATAEAFDEBGMWRTGD--VAVVDSGGMHR--IYGRSVD 392
QY 443 FKLSGIVFVSGPLRTAVLEGSSVYLDV--VVAAPRECLGLVFPRLDORALSGLGE 501
DB 393 LKLGGRIGAGET--EMALLGHPDVREAVVGLPD--EDLQGRIVAFV-----VGAE 441
QY 502 ASDAEVLAS-----EPVAMPADMLK 523
DB 442 ALDADELINVAQQLSIHKRPREVRFDALPR 473

RESULT 10
ID ADL23382 standard; protein; 700 AA.
AC ADL23382;
XX

DT	20-MAY-2004	(first entry)	
XX			
XX	Plant acyl-coenzyme A synthetase amino acid sequence #6.		
XX			
KW	plant; acyl-coenzyme A synthetase; acyl-CoA synthetase; enzyme;		
KW	transgenic plant.		
XX			
OS	Unidentified.		
PN	WO200209295-A2.		
XX			
PD	31-JAN-2002.		
XX			
PF	19-JUL-2001; 2001WO-US0222774.		
XX			
PR	21-JUL-2000; 2000US-0220474P.		
XX			
PA	16-JUL-2001; 2001US-00906419.		
PA	(SHOC/) SHOCKEY J M.		
PA	(SCHN/) SCHNUR J.		
PA	(BROW/) BROWSE J A.		
PI			
XX	Shockey JM, Schnurr J, Browse JA;		
DR	WPI, 2002-241594/29.		
XX			
XX	N-PSDB; ADL23371.		
PT			
PT	Novel acyl coenzyme A synthetases gene useful for altering a phenotype of		
PT	a plant, making a transgenic plant and for producing variants of acyl-CoA		
PT	synthetases.		
XX			
PS	Claim 2; SEQ ID NO 17; 155pp; English.		
XX			
CC	The invention comprises the amino acid and coding sequences of plant acyl		
CC	-coenzyme A synthetase (acyl-CoA synthetase) enzymes. The DNA and protein		
CC	sequences of the invention are useful for altering a phenotype of a plant		
CC	(transgenic plant). The DNA and protein sequences of the invention are		
CC	also useful for producing variants of acyl-CoA synthetases. The present		
CC	amino acid sequence represents a plant acyl-CoA synthetase of the		
CC	invention.		
XX			
XX			
SQ	Sequence 700 AA;		
	Query Match	9.2%; Score 283; DB 5; Length 700;	
	Best Local Similarity	22.5%; Pred. No. 5,1e-18;	
	Matches 150; Conservative 97; Mismatches 214; Indels 206; Gaps 35;		
QY	9 PEPGILRLERLHMATREPOTCVARRAN-----GEMRKISYEMHNRAINQSLIPYGL 64		
DB	81 PEIGLHNFVHAVETVAENKYLGRVSDGTIGBSWMTYGEAASEQALIGSGLLFHGV 140		
QY	65 S-----AERPLLIIVSGNDLEHLQAFGAMVAGIFPCVPSPAYSLSLOPLAK--LRH 113		
DB	141 NQGDCVGLYFINRPEWLV-----VDIACAAYISVSV-----PLYDLGPAAVVFVNVH 188		
QY	114 -----IVGLLOGLVFAADADAFORAIEI--LPDDVAIFTRGELAGRRVSPSL 163		
DB	189 ANLQAFVCPQTLNILLSFLAE-IPSIRLIVVVGADDEHLPSL-PRG--TGVTIVSYQKL 244		
QY	164 LEQPGGIIADNNAFAATPDTIAKLFETGSGTGLPAAVPTTQMLCAN--QOMLIQTPV 220		
DB	245 LSQ--GRSSLHFSPKPEDIATICTYTGTTPKGVVLTHNMLANVAGSSVEAEFFPP- 301		
QY	221 FGEEDPVAVDMLPMNH-----TGGSGHNIGVLYNGSTYYLDDQKPAQQAFAETL 270		
DB	302 ----SDVYISYFLPHAIYERANQINGVYG--VAVGFYQGVFLKMD----- 342		
QY	271 NNLSEISPTAYLTVPKGWEELVGALEKSDTLREFFAAMKLFPPAAAGLSQGI----- 323		
DB	343 -DFAYLARTPIFGSVRLNRIYDGI--TSAVASSGVVKKRLFEIAYNSKKQAINGRTPS 399		
QY	324 --WDL--DRVAQHCGERIRMA-----GLGTEFAPSCT 355		

Db	400	AFWMDLVENKIKKEK-LGGRVRFMSGASPLSPDVMDFLRICPGCSVRBEYGMTET--SCV	455
Qy	356	FTT---GFLSMAGYIGLPAPEGCEVTLVPD-----GKLEGFHFQPHVMSGYW	399
Db	457	ISAMDDGD-NLSGHVSGSNPACSEVTLVVPPEMAYNTSDQPYRGEICVR--GPILFGKY	513
Qy	400	RAPEQNAAPFEEGGYVCGSDAIKLADPADDPQKGLMFDG-----RIADPFLSSGVFOS	452
Db	514	KDEBETREILDDGDLHTGDTI-----GLWLPGRGLKTIIDKKNIIFKLAGEYTA	562
Qy	453	VGPLR-----TRAVLEGGSY---VLDVVVAAPD-----RECLGLLV-PP	487
Db	553	PEKIEINYTKCRFVSQCIFIHGSPFNSLVAIYSVDPPEVMKMDAASBGIKYEHLGQCNPD	622
Qy	488	RLLD--CRALSGLGEA-----SDAEVLASEP-----VRAMFADW	520
Db	623	RVRKTVLAEMDDIGREAOLRGFEEFAKAVTLVPEPFTLENGLLTPTPFKIKRPOAKAYFAEA	682
Qy	521	LKRLNRE	527
Db	683	ISKMYAE	689
RESULT 11			
ID	ADL72359	standard; protein; 700 AA.	
AC	ADL72359;		
DT	20-MAY-2004	(first entry)	
DE	Modified Arabidopsis thaliana acyl-CoA synthetase 3B protein.		
KW	plant; acyl-CoA synthetase; soybean; sunflower; cotton; maize; castor; transgenic plant; triacylglycerol biosynthesis; fatty acid; seedling; beta-oxidation cycle substrate; jasmonic acid; plant defence.		
OS	Arabidopsis thaliana.		
PN	WO2003087321-A2.		
PD	23-OCT-2003.		
PF	09-APR-2003; 2003WO-US010754.		
PR	09-APR-2002; 2002US-00119136.		
PR	08-APR-2003; 2003US-00119136.		
PA	(UTWA-) UNIV WASHINGTON STATE RES FOUND.		
PI	Shockey JM, Schnurr J, Browse JA;		
DR	WPI: 2003-853948/79. N-PSDB; ADL72353.		
PT	New plant acyl-CoA synthetase protein derived from soybean, sunflower, cotton, maize, and castor, useful in cuticle was synthesis, and in the synthesis of jasmonic acid which is involved in reproduction and plant defense.		
PS	Claim 29; SEQ ID NO 130; 226pp; English.		
CC	The invention relates to a new purified plant acyl-CoA synthetase protein comprising at least one of the motifs selected from 9 fully defined motifs given in the specification, and derived from a crop plant selected from soybean, sunflower, cotton, maize, and castor. The purified plant acyl-CoA synthetase protein comprises at least one motif selected from: V-P/T-L/I-Y-D/A/S-T/S-L-G; I-M-C-Y/F/K-T-S-G-T/S-T/S-G-X1-P-K-G-V; S/A-Y/M/F-P-L/S-A/M-H; L/O-K/R-P-T/P/S; S/G/V/-G-A/G/S-N/L/S-P-L/I/M; G-Y-G-L/M-T-E-T/S; P/S/A-R/K-G-A-B/I-V-C/K/V-I/V/L-R-G-G; IDRRK; and L-V/L/M/L-P/A-T/A/S-F/L/M/Y-K-X1-K/R-R. The nucleic acid is useful in producing a transgenic plant (claimed). The plant Acyl-CoA synthetase is useful in TAG biosynthesis, for activating fatty acids released from oil		

CC bodies in newly germinated seedlings, as substrates for the beta-oxidation cycle which supplies the plant with cellular energy until it becomes photosynthetically competent, in cuticle was synthesizing, and in the synthesis of jasmonic acid, a fatty acid-derived signaling compound involved in reproduction, plant defence, and other response reactions. CC This sequence corresponds to a modified ACS3B protein of the invention.

XX Sequence 700 AA:

Query Match 9.2%; Score 283; DB 7; Length 700;
Best Local Similarity 22.5%; Pred. No. 5.1e-18;
Matches 150; Conservative 97; Mismatches 214; Indels 206; Gaps 35;

```

QY 9 PPGRIERLEHNAKTRPEQTCVAARAAN-----GEMRISYAEFMHVRALIAQSLPYGL 64
DB 81 PEIGTLHDNFVAHYEAENKYLGTNRSDGTIGESWMYTGGAASRQAIIGSGLPHGV 140
QY 65 S-----AERPLIVSGNDLEHLQAFGAMYAGIPYCPSPAYSLLSQDLAK--LRH 113
DB 141 NQGDGCVGLYFIRNPEMLV-----VDHCAAYSFVS-----PLYDTLGPDAVKPFVNH 188
QY 114 -----IVGLIQGLVPAADAAPFORAETI--LPDDVPALFTGELAGRTVSFDSL 163
DB 189 ANLQALFCVPQTINILISFLAE--IPSRILVVGADENHPSL--PG--TGVTVSYOKL 244
QY 164 LEQPGIEADNAPATGPDITIAKFLFTSGSTKLPKAVPTTORMICAN---QOMLQTFPV 220
DB 245 LSG--GRSSLHPPSPKPEBIDATICTSGTTPKGVLTGMLIANVAGSSVEAEFFP-- 301
QY 221 FGSEPPVLVDMLEPMNH-----TFGSHNIGIYLVNGTGYLDDGKPTAQGFATL 270
DB 302 ----SDVYISYLPDAHIERANOIMGVYGG---VAVGFYQGDVFKLMD----- 342
QY 271 RNLSISPTAYLVTPKGMEEVLGALERTSLRERFARMKLPFPAAAGLSQGI----- 323
DB 343 -DFAVLRPTIFCSPVRLYNYIDGI--TSAVKSQGVKKLFEIAYNSKKQAIINGRTPS 359
QY 324 --WDRL--DRVAHCHGERIRMA-----GIMTETAPSC 355
DB 400 AFMDKLVFNKIKER--LQGRVRFMGSGASPLSPVDMFLRICGCSVREGYGMET--SCV 456
QY 356 FTN---GPLSMAGYIGLPARCEVKLPVVD-----GKLEGRFHGPHVNSGYW 359
DB 457 ISMDGDD-NLSGHWSPNACEVKLVDPENMYTSDQYPRGELCVR--GPIIFGY 513
QY 400 RAEQNAQAEDDEGYCCSDAIIKLADPADPOKIMFDG-----RIADPKLSSGYFVS 452
DB 514 KDEEQTRREILDGDMHTTGD-----GLMLFGRLKIIRKKNIFKLAQGEYIA 562
QY 453 VGPLR-----TRAVLEGGSY--VLDVVVAAP-----RECGLLV-FP 487
DB 563 PEKIEENVYTKCRFVSOCTFHGDSFNSSLVAIVSDVPEVMKMAASGIKYEHIGQLNDP 622
QY 488 RLTD--CRALSGIKEA-----SDAEVLASEP-----VRAMADW 520
DB 623 RVAKTYLAEMDDIGREHQALGFEPAAKAVTLVPERFTLENGLLPTFKIKRPAKAYRAEA 682
QY 521 LKRLNR 527
DB 683 ISKRYAB 689

```

RESULT 12
ADN72469
ID ADN72469 standard; protein; 700 AA.
XX
XX ADN72469;
XX
XX 15-JUL-2004 (first entry)
XX Thale cress protein upregulated in E2fa/Dpa expressing plants Seqid 364.
XX plant; transgenic; E2fa/Dpa transcription factor; growth regulator;
XX

KM animal feed product; thale cress; cell wall biosynthesis;
KM nitrogen metabolism; carbon metabolism.

XX Arabidopsis thaliana.

XX WO2004035798-A2.

XX 29-APR-2004.

XX 20-OCT-2003; 2003WO-BP011658.

XX 18-OCT-2002; 2002EP-00079408.

XX (CROP-) CROPDISEIGN NV.

XX Inze D, De Veylder L, Vlieghe K;

XX WPI; 2004-348466/32.

XX N-PSDB; ADN72468.

PT Altering plant characteristics, useful for producing plants for enzyme or pharmaceutical production comprises modifying in a plant, expression of one or more nucleic acids and/or modifying level or activity of one or more proteins.

PS Claim 1; SEQ ID NO 364; 134pp; English.

CC This invention relates to a novel method for altering one or more plant characteristic. Specifically, it refers to identifying genes that are up- or down-regulated in transgenic plants overexpressing the heterodimeric E2fa/Dpa transcription factor of Arabidopsis and using these sequences to alter plant characteristics accordingly. The present invention describes generating transgenic plants for the production of growth regulators, enzymes, therapeutics, pharmaceuticals and animal feed products, where the altered plant characteristics are selected from increased yield or biomass, enhanced survival capacity, stress tolerance, plant architecture or physiology, altered endoreplication, biochemistry, signal transduction, storage lipid mobilization and/or altered photosynthesis, each relative to the corresponding wild type plants. Accordingly, these sequences can also be useful as positive or negative selectable markers during transformation of cells or tissues. The identified genes play a role in a variety of biological processes such as DNA replication, cell wall biosynthesis, nitrogen and/or carbon metabolism or they function as transcription factors. This polypeptide sequence is thale cress protein expressed by a gene upregulated 1.3 fold or more in plants overexpressing the E2fa/Dpa transcription factor, given in an exemplification of the invention.

XX Sequence 700 AA:

Query Match 9.2%; Score 283; DB 8; Length 700;
Best Local Similarity 22.5%; Pred. No. 5.1e-18;
Matches 150; Conservative 97; Mismatches 214; Indels 206; Gaps 35;

```

QY 9 PPGRIERLEHNAKTRPEQTCVAARAAN-----GEMRISYAEFMHVRALIAQSLPYGL 64
DB 81 PEIGTLHDNFVAHYEAENKYLGTNRSDGTIGESWMYTGGAASRQAIIGSGLPHGV 140
QY 65 S-----AERPLIVSGNDLEHLQAFGAMYAGIPYCPSPAYSLLSQDLAK--LRH 113
DB 141 NQGDGCVGLYFIRNPEMLV-----VDHCAAYSFVS-----PLYDTLGPDAVKPFVNH 188
QY 114 -----IVGLIQGLVPAADAAPFORAETI--LPDDVPALFTGELAGRTVSFDSL 163
DB 189 ANLQALFCVPQTINILISFLAE--IPSRILVVGADENHPSL--PG--TGVTVSYOKL 244
QY 164 LEQPGIEADNAPATGPDITIAKFLFTSGSTKLPKAVPTTORMICAN---QOMLQTFPV 220
DB 245 LSG--GRSSLHPPSPKPEBIDATICTSGTTPKGVLTGMLIANVAGSSVEAEFFP-- 301
QY 221 FGSEPPVLVDMLEPMNH-----TFGSHNIGIYLVNGTGYLDDGKPTAQGFATL 270
DB 302 ----SDVYISYLPDAHIERANOIMGVYGG---VAVGFYQGDVFKLMD----- 342

```

QY		271	RNLSEISPTAVLTVPKGMELVGALERDSTLERFFPAKMKLPEFPAAAGLSOGI-----	323
Db		343	-DPAVLRFETICSVRLNRIIDGT--TSVAVSGGVKKGLFEIAYNKKQALINRQT8	399
QY		324	--WDRL--DRVAEOHCHGERIRMA-----GIGMTETAASCT	355
Db		400	AFWMDPLVFNTIKER-LGGRRFRFMGSASPSPDPVMDFLRICGCSVREGVMGTET--SCV	456
QY		356	FTT---GELSMAGYIGLAPGCEVTLVPD-----GKLEGRFHGHVMSGW	399
Db		457	ISAMDDGG-NISGHGSSPNPACAEVTLVDPENNYTSDDPYPRGELCVR--GPLIFKGY	513
QY		400	RAPENAOAPEDEEGYCCGDATIKLADPADPOKGLMPDG-----RIADPFKSSGVFVS	452
Db		514	KDEOQTRITLDGCGMLHTGDI-----GLWLPGSRLLTIIDRKNIIFKLAGEEYIA	562
QY		453	VGPRLR-----TRAVLGGSY--VLDVVVAAPD-----RECTGLLV-PP	487
Db		563	PEKIENVYTCKRFVSGCFIHGDSFNSSLVAIVSVPBEVKDMAASBGIKYEHLGQLCNDP	622
QY		488	RLLD--CALSGIKGEA-----SDAEVLASEP-----VRAMPADM	520
Db		623	RVRKTVALAMDOLGREAQLRGFEPFAKAVTLVEBPFTLENGLLTPTFKIRPKAKAYFAEA	682
QY		521	LKRLNRE	527
Db		683	ISKMYAE	689
RESULT 13				
ADL72422	ID	ADL72422	standard; peptide; 700 AA.	
XX	ADL72422;			
AC	ADL72422;			
XX	20-MAY-2004 (first entry)			
DT				
XX	Arabidopsis thaliana ACS derived peptide #9.			
DE				
XX	plant; acyl-CoA synthetase; soybean; sunflower; cotton; maize; castor;			
KM	transgenic plant; triacylglycerol biosynthesis; fatty acid; seedling;			
KW	beta-oxidation cycle substrate; jasmonic acid; plant defence.			
XX				
OS	Arabidopsis thaliana.			
XX				
PN	WO2003087321-A2.			
PD				
XX	23-OCT-2003.			
XX				
PF	09-APR-2003; 2003WO-US010754.			
PR				
PR	09-APR-2002; 2002US-00119136.			
PR	08-APR-2003; 2003US-00119136.			
XX				
PA	(UYWA-) UNIV WASHINGTON STATE RES FOUND.			
XX				
PI	Shockey JM, Schnurr J, Browse JA;			
XX				
DR	WPI; 2003-853948/79.			
XX				
PT	New plant acyl-CoA synthetase protein derived from soybean, sunflower,			
PT	cotton, maize, and castor, useful in cuticle waxy synthesis, and in the			
PT	synthesis of jasmonic acid which is involved in reproduction and plant			
PT	defence.			
XX				
PS	Disclosure; SEQ ID NO 208; 226pp; English.			
XX				
CC	The invention relates to a new purified plant acyl-CoA synthetase protein			
CC	comprising at least one of the motifs selected from 9 fully defined			
CC	motifs given in the specification, and derived from a crop plant selected			
CC	from soybean, sunflower, cotton, maize, and castor. The purified plant			
CC	acyl-CoA synthetase protein comprises at least one motif selected from: V			

Query Match	9.1%	Score 282;	DB 7;	Length 700;
Best Local Similarity	22.5%	Prod. No. 6,4e-16;		
Matches 150;	Conservative 96;	Mismatches 215;	Indels 206;	Gaps 35;
Qy	9	PPGRIRLERLEHMAKTRPEQTCVAAARAAN----	GEMRIRISAEEMFNHVRALASGLPLPGL	64
Db	81	PEITGLINDNFHAVETIAENKYLQTRVRSDDTIEYMNTYGEAASEKQALIGSGLLFFRGV		140
Qy	65	S-----AERPLLIVSGNDLEHLOAFGAMYAGIIPYCVSPAYSILLSODLAK--LRH		113
Db	141	NOGDCVGLYFIRNREMYL-----VDHAAAYSFVSF-----FLYDTLGDPAVYFVNH		188
Qy	114	-----IYGLDQGLVPAADANPFORAIETI--LPDDVAIPIFRGLAARRYSPSL		163
Db	189	ANLOAICFCVQTLNILLSFLAE--IPSRLLIYVGGADENHPSL--PRG--TGVTVISYQKL		244
Qy	164	LEDPGEGIADNAFAATGPDTIAKELFTSGSTKLEKAVFTQRMLCAN--QOMLLQTFPV		220
Db	245	LSQ--GRSSLHPFPSPREDIATICTYSGTGTGPKGVLTIGNLIANVAGSSVAEFPFP--		301
Qy	221	FGEERPVLVDMLEPMNH-----TFGSHNIGIYLVNGSGTYLDDCKRTAGFAETL		270
Db	302	-----SDVYISLPLPAHYERANQIMYVGG--VAVGFIGQDVRLKMD-----		342
Qy	271	RNLSEISPTAVLIVPKGMEELVGLERDSTLERFFARMKLFPPAAGLSOGI-----		323
Db	343	-DFAVLAPFTIFCSVPLRYLNRIYDGI--TSAYKSGGVKKRLFEIAYNSKKQALINGRTPS		399
Qy	324	--WDRL--DRVAEOHCERIRMAA-----	GLGNTETAPBCT	355
Db	400	AFMFKLIVFNKIKKK--LGKRVAFPMKSGASPLSPDVWDFLRICFGSVBEGYAMTET--SCV		456
Qy	356	FTT-----PLSMAGYIGLPAQCEYKLVVD-----	GKLEGRFPGPNVMSGYM	399
Db	457	ISAMDDGD--NLSGAVGSPNPAACEKLVVDPEMNTSDQPIRBEICKR--GPIIIFKGY		513
Qy	400	RAPRONAOPDEEGYVSGDAIKLADPADPQGLMPDG-----	RIADEFKLSSGVFS	452
Db	514	KDEQOTREIILDGDXLHTGDI-----	GLMXGGRUKITDRKKNIIFKLAAQGEYIA	562
Qy	453	VGPRL-----TRAVLEGSTV--VLQVYVAAPD-----	RECLGLIV--FP	487
Db	563	PEKLENNYTKCRFVSGCFIHGDSFNSSLVALVSDPEVMKDMAASEGITYENHLQLCNDR		622
Qy	488	RLLD--CRALSGLGKEA-----SDAEVLASEP-----	VRAMPADW	520
Db	623	RVKRTVLAEMDDGLQREQLRGFEAKAVTLVPEPFTLENGLLTPTFKIKRQAQAAVAFEA		682
Qy	521	LKRLIRE	527	
Db	683	ISKMYAE	689	

RESULT 14

ADL72260

ID ADL72260 standard; protein; 700 AA.

AC ADL72260;
 XX
 DT 20-MAY-2004 (first entry)
 XX
 DE Arabidopsis thaliana acyl-CoA synthetase 3B protein.
 XX
 KM plant; acyl-CoA synthetase; soybean; sunflower; cotton; maize; castor;
 KM transgenic plant; triacylglycerol biosynthesis; fatty acid; seedling;
 KM beta-oxidation cycle substrate; jasmonic acid; plant defence.
 XX
 OS Arabidopsis thaliana.
 PN W02003087321-A2.
 XX
 PD 23-OCT-2003.
 XX
 PF 09-APR-2003; 2003WO-US010754.
 XX
 PR 09-APR-2002; 2002US-00119136.
 PR 08-APR-2003; 2003US-00119136.
 XX
 PA (UYMA-) UNIV WASHINGTON STATE RES FOUND.
 PI Shockey JM, Schnurr J, Browne JA;
 XX
 DR WPI: 2003-853948/79.
 DR N-PSDB: ADL72249.
 XX
 PT New plant acyl-CoA synthetase protein derived from soybean, sunflower,
 PT cotton, maize, and castor, useful in cuticle was synthesis, and in the
 PT synthesis of jasmonic acid which is involved in reproduction and plant
 PT defense.
 XX
 PS Disclosure; SEQ ID NO 17; 226pp; English.
 XX
 CC The invention relates to a new purified plant acyl-CoA synthetase protein
 CC comprising at least one of the motifs selected from 9 fully defined
 CC motifs given in the specification, and derived from a crop plant selected
 CC from soybean, sunflower, cotton, maize, and castor. The purified plant
 CC acyl-CoA synthetase protein comprises at least one motif selected from: V
 CC -P/T-L-I-Y-D/A/S-T/S-L-G; I-M-C-Y/F/K-T-S-G-T/S-T/S-G-XI-P-K-G-V; S/A-
 CC -Y/M-F-L-P-L-S-A/W-H; L/Q-K/R-P-T/P/S; S/G/V/-G-A/G/S-A/L/S-P-L/I/W; G-Y-G
 CC -L/M-T-E-T/S; P/S-A/R-K-G/A-E/I-V-C/K/V-I/V/L-R-G-G; IIPKK; and L-
 CC -L/V/M/L-P-A-T/A/S-F/L/M/Y-K-XI-K/R-R. The nucleic acid is useful in
 CC producing a transgenic plant (claimed). The plant acyl-CoA synthetase is
 CC useful in TAG biosynthesis, for activating fatty acids released from oil
 CC bodies in newly germinated seedlings, as substrates for the beta-
 CC oxidation cycle which supplies the plant with cellular energy until it
 CC becomes photosynthetically competent. In cuticle was synthesis, and in
 CC the synthesis of jasmonic acid, a fatty acid-derived signaling compound
 CC involved in reproduction, plant defence, and other response reactions.
 CC This sequence corresponds to an ACS protein of the invention.
 CC
 SQ Sequence 700 AA;
 Query Match 9.1%; Score 279; DB 7; Length 700;
 Best Local Similarity 22.7%; Pred. No. 1.3e-17;
 Matches 150; Conservative 100; Mismatches 216; Indels 196; Gaps 35;
 QY 9 PFGRIILERLEHMAKTRPEQTCVAAARAAN---GEMRISYAEMFHNVRATAGSLPYGL 64
 DB 81 PEIGTLHDNFVHAYETAEKNYLGTRVRSDDTIEYGMWMTGEVAAERQAGSLDTHGV 140
 QY 65 S-----AERLLVSGNDEHLQALAGAMTAGIPYCVSPAYSLISODLAK--LRH 113
 DB 141 NOGACVGLYFLNREMLV---VDHACAAVSFVS-----PLYDTLGDAVKFVVMH 188
 QY 114 -----IYGLQPGLVFAADAPFORAETI--LPDDVPAITRSGLAGRRVSPDSL 163
 DB 189 ANLOATFCVQPTLILSLFAE-IPSTRLLVVGADENHLSL-PRG--TGVTVVSTQKL 244
 QY 164 LEOPGGIEADNAPATGPDITAKFLFTSGSTKLPRKAVPTTQRMCAN---QOMLLQTFPV 220

DB 245 LSG--GRSSLHPSPPKPEDIATICYTSGTTPKGVVLTGNLIANVAGSSVEAEFRP- 301
 QY 221 FGEPPPLVLDMLPMNHPFGGSHNIGVLYNGCTYYLDDGKPTAOGF-----AEILRNISE 275
 DB 302 ----SDVIISYLPPLAHAYERANOI-MGVYGG-----VAAGFYQGDVLKLMDDPAV 346
 QY 276 ISPTAVLTVPKGMDELVALERDSTLRERFPABMKLFFFAAAGLSOGI-----WDR 326
 DB 347 LRPTIFGCVPRVLYRITDGI--TSAVKSSGVKKRLERIAVNSKKOALINRTSAPFMDK 404
 QY 327 L--DRVAEQHCEGIRIEMA-----GLGMTETAPSCFTTT-- 358
 DB 405 LVFNKIKER--LGVVRFMGSGASPLSPDVMDFLRICGCSVREGYGTET--SCVISAMD 461
 QY 359 -GPLSMAGYIGLPAFGCEVULVPD-----GKLEGFPHGPHVSGVWRAPQ 404
 DB 462 DGD-NLSGHVGSFPNACEVXLVDPEMNYSSEDPYPRGELCYR--GPILFKGYKDEQ 518
 QY 405 NQAQFDEGYGCSGDALKLADPADPQKGLMPDG-----RIADPFLKSGVFSVGPLR 457
 DB 519 TREILDDGMLHTGDI-----GLWLPGRRLKTIIDRKXIFFLAQSXYIAPEKIE 567
 QY 458 -----TRAVLEGSY--VLDVVVAAPD-----RECLGLLV-PPTLID- 491
 DB 568 NVYTKCRFVSQCFIHGDSFNSLVAIVSDPEVKKDWAASBGIKYEHLGQCNDRVRKT 627
 QY 492 -CRALSGLGEKA-----SDAEVLASEP-----VRANFAMLKRLN 525
 DB 628 VLAEMLDGLRBAQURGEFPAKAVTLVPEPTLENGLLTPPFKIRPOAKVFAEAIKMY 687
 QY 526 RE 527
 DB 688 AE 689
 RESULT 15
 ABP6163
 ID ABP6163 standard; protein; 695 AA.
 AC
 AC ABP6163;
 DT 19-NOV-2002 (first entry)
 XX
 DE Bifidobacterium longum NCC2705 ORF amino acid sequence SEQ ID NO:907.
 XX
 KM Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;
 KM anti-diarrhetic; antibacterial; inhibitor of Salmonella; detection;
 KM identification; lactic acid bacterium; diarrhoea; pathogenic bacteria;
 KM rotavirus; food composition; pharmaceutical composition.
 XX
 OS Bifidobacterium longum.
 XX
 PN EP1227152-A1.
 XX
 PD 31-JUL-2002.
 XX
 PF 30-JAN-2001; 2001EP-00102050.
 PR 30-JAN-2001; 2001EP-00102050.
 XX
 PA (NEST) SOC PROD NESTLE SA.
 XX
 DR WPI: 2002-668397/72.
 XX
 PT Novel polynucleotide comprising Bifidobacterium genome sequence useful as
 PT a probe or primer for detecting and/or identifying Bifidobacterium longum
 PT in a biological sample.
 XX
 PS Claim 3; SEQ ID NO 907; 80pp; English.
 CC The present invention describes a polynucleotide (I) comprising a
 CC sequence of a Bifidobacterium genome selected from the nucleotide
 CC sequences given in AB081842 and AB081843, or a sequence exhibiting at

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2005, 23:47:50 ; Search time 105 Seconds

(without alignments)
2026.661 Million cell updates/sec

Title: US-09-750-986D-30

Perfect score: 3082

Sequence: 1 MRLEALPRGRILRLH.....KVDALYGEQSMRLDEATL 589

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1608061 seqs, 361289386 residues

Total number of hits satisfying chosen parameters: 1608061

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/prodata/2/pubppaa/US07_PUBCOMB.pep:*

2: /cgn2_6/prodata/2/pubppaa/PCT_NEW_PUB.pep:*

3: /cgn2_6/prodata/2/pubppaa/US06_NEW_PUB.pep:*

4: /cgn2_6/prodata/2/pubppaa/US06_PUBCOMB.pep:*

5: /cgn2_6/prodata/2/pubppaa/US07_NEW_PUB.pep:*

6: /cgn2_6/prodata/2/pubppaa/PCTUS_PUBCOMB.pep:*

7: /cgn2_6/prodata/2/pubppaa/US08_NEW_PUB.pep:*

8: /cgn2_6/prodata/2/pubppaa/US08_PUBCOMB.pep:*

9: /cgn2_6/prodata/2/pubppaa/US09_PUBCOMB.pep:*

10: /cgn2_6/prodata/2/pubppaa/US09B_PUBCOMB.pep:*

11: /cgn2_6/prodata/2/pubppaa/US09C_PUBCOMB.pep:*

12: /cgn2_6/prodata/2/pubppaa/US09_NEW_PUB.pep:*

13: /cgn2_6/prodata/2/pubppaa/US10_PUBCOMB.pep:*

14: /cgn2_6/prodata/2/pubppaa/US10B_PUBCOMB.pep:*

15: /cgn2_6/prodata/2/pubppaa/US10C_PUBCOMB.pep:*

16: /cgn2_6/prodata/2/pubppaa/US10D_PUBCOMB.pep:*

17: /cgn2_6/prodata/2/pubppaa/US10E_NEW_PUB.pep:*

18: /cgn2_6/prodata/2/pubppaa/US11_NEW_PUB.pep:*

19: /cgn2_6/prodata/2/pubppaa/US60_NEW_PUB.pep:*

20: /cgn2_6/prodata/2/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	3082	100.0	589 8 US-08-976-063C-30	Sequence 30, Appl
2	3082	100.0	589 10 US-09-750-986D-30	Sequence 20490, A
3	1457	47.3	575 14 US-10-369-493-20490	Sequence 11697, A
4	1294	42.0	576 14 US-10-369-493-11697	Sequence 14740, A
5	1294	42.0	576 14 US-10-369-493-14740	Sequence 15216, A
6	1294	42.0	576 14 US-10-369-493-15216	Sequence 14320, A
7	1250	40.6	553 14 US-10-369-493-14320	Sequence 17047, A
8	972	31.5	621 14 US-10-369-493-17047	Sequence 244, App
9	383.5	12.4	544 14 US-10-369-493-244	Sequence 10394, A
10	333	10.8	636 14 US-10-156-761-13094	Sequence 10386, A
11	324	10.5	582 14 US-10-369-493-10386	Sequence 17, Appl
12	283	9.2	700 10 US-09-906-419-17	Sequence 130, App
13	283	9.2	700 14 US-10-119-136-130	

14	283	9.2	700 15 US-10-410-031-130	Sequence 130, App
15	281	9.1	585 14 US-10-369-493-311	Sequence 311, App
16	281	9.1	633 14 US-10-369-493-19735	Sequence 19735, A
17	279	9.1	700 14 US-10-119-136-17	Sequence 17, Appl
18	279	9.1	700 15 US-10-410-031-17	Sequence 17, Appl
19	278.5	9.0	613 14 US-10-369-493-20886	Sequence 20886, A
20	278	9.0	562 14 US-10-369-493-13736	Sequence 13736, A
21	278	9.0	701 10 US-09-906-419-16	Sequence 16, Appl
22	278	9.0	701 15 US-10-119-136-16	Sequence 16, Appl
23	278	9.0	701 15 US-10-410-031-16	Sequence 16, Appl
24	277.5	9.0	504 14 US-10-369-493-12089	Sequence 12089, A
25	277.5	9.0	598 14 US-10-369-493-8109	Sequence 8109, App
26	273	8.9	677 17 US-10-739-930-6854	Sequence 6854, App
27	272	8.8	589 14 US-10-369-493-21245	Sequence 21245, A
28	268	8.7	489 14 US-10-369-493-20547	Sequence 20547, A
29	268	8.7	500 14 US-10-369-493-19511	Sequence 19511, A
30	265	8.6	601 14 US-10-369-493-923	Sequence 923, App
31	265	8.6	638 14 US-10-156-761-12352	Sequence 12352, A
32	265	8.6	6842 15 US-10-461-194-131	Sequence 131, App
33	264.5	8.6	546 10 US-09-838-469-23	Sequence 23, Appl
34	263.5	8.5	544 14 US-10-378-168-23	Sequence 23, Appl
35	263	8.5	634 15 US-10-713-810-3	Sequence 3, Appl
36	263	8.5	696 17 US-10-425-115-300648	Sequence 300648, Sequence 165496,
37	263	8.5	858 16 US-10-437-963-165496	Sequence 24, Appl
38	262.5	8.5	544 10 US-09-838-469-24	Sequence 2, Appl
39	262.5	8.5	544 10 US-09-813-2798-4	Sequence 24, Appl
40	262.5	8.5	544 14 US-10-378-168-24	Sequence 24, Appl
41	262.5	8.5	544 14 US-10-378-168-45	Sequence 45, Appl
42	262.5	8.5	544 16 US-10-655-878-2	Sequence 2, Appl
43	262.5	8.5	544 16 US-10-655-878-4	Sequence 4, Appl
44	262.5	8.5	544 16 US-10-655-878-4	Sequence 67862, A
45	262.5	8.5	4317 15 US-10-282-122A-67862	

ALIGNMENTS

RESULT 1

US-08-976-063C-30

Sequence 30, Application US/08976063C

Publication No. US20020182697A1

GENERAL INFORMATION:

APPLICANT: Alexander Steinhuechel; Horst Priefert; Jurgen Rabenhorst

TITLE OF INVENTION: SYNTHETIC ENZYMES FOR THE PRODUCTION OF

TITLE OF INVENTION: CONFERTIL ALCOHOL, CONFERTILALDEHYDE, FERULIC ACID, VANILLIN AL

TITLE OF INVENTION: ACID AND THEIR USE

NUMBER OF SEQUENCES: 42

CORRESPONDENCE ADDRESSES:

ADDRESSEE: SPRUNG KRAMER SCHAEFER & BRISCOE

STREET: 660 White Plains Road

CITY: Tarrytown

STATE: New York

COUNTRY: U.S.A.

ZIP: 10591-5144

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB storage

COMPUTER: HP VECTRA

OPERATING SYSTEM: DOS

SOFTWARE: WordPerfect 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/976, 063C

FILING DATE: 21-NOV-1997

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 196 49 655.1 (Germany)

FILING DATE: 29-NOV-1996

ATTORNEY/AGENT INFORMATION:

NAME: Kurt G. Briscoe

REGISTRATION NUMBER: 33,141

REFERENCE/DOCKET NUMBER: Bayer 9998-CAO

TELECOMMUNICATION INFORMATION:

TELEPHONE: (914) 332-1700

TELEFAX: (914) 332-1844

TELEX:
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 589 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-976-063C-30

Query Match 100.0%; Score 3082; DB 8; Length 589;
Best Local Similarity 100.0%; Pred. No. 7.7e-289;
Matches 589; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MRSLEALPPPGILRLERLHMAKTRPEQTCVAAARAANGEMRRISYAEFHNVAIAQSLL 60
DB 1 MRSLEALPPPGILRLERLHMAKTRPEQTCVAAARAANGEMRRISYAEFHNVAIAQSLL 60
QY 61 PYGLSARPLLIVSGNDLEHLQLAFGAMYAGIPYCVSPAYSLSDLAKLRHIVGLLP 120
DB 61 PYGLSARPLLIVSGNDLEHLQLAFGAMYAGIPYCVSPAYSLSDLAKLRHIVGLLP 120
QY 121 GLVFAADAAFPORAETIIPDDVPALFTRGELAGRTVSFDSLLEPGGIEADNAFAATG 180
DB 121 GLVFAADAAFPORAETIIPDDVPALFTRGELAGRTVSFDSLLEPGGIEADNAFAATG 180
QY 181 PDTIAKFLFTSGSTKLPAKAVPTTQRMLCANQOMLQTFPVFGSEPPVLDMLPMNHTFGG 240
DB 181 PDTIAKFLFTSGSTKLPAKAVPTTQRMLCANQOMLQTFPVFGSEPPVLDMLPMNHTFGG 240
QY 241 SHNIGIVLVNGGTYLDDGKPTAQGFATLRNLSEISPTAYLTVPKGMEELVGALERDST 300
DB 241 SHNIGIVLVNGGTYLDDGKPTAQGFATLRNLSEISPTAYLTVPKGMEELVGALERDST 300
QY 301 LRERFPAARMKLFPPAAAGISOGIMDRLDRAVEOHCGERIRMMAGLQMTETAPSCFTTGP 360
DB 301 LRERFPAARMKLFPPAAAGISOGIMDRLDRAVEOHCGERIRMMAGLQMTETAPSCFTTGP 360
QY 361 LSMAGYIGLPAPCEVTLVVDGKLEGRFHGPHVMSGYRPAEONQAQDEEGYCSGDA 420
DB 361 LSMAGYIGLPAPCEVTLVVDGKLEGRFHGPHVMSGYRPAEONQAQDEEGYCSGDA 420
QY 421 IKLADPADPQKGMFDGRIAEDEKLSGVSFVSGPLRTAIVLEGGSYLDVVVAADPRRC 480
DB 421 IKLADPADPQKGMFDGRIAEDEKLSGVSFVSGPLRTAIVLEGGSYLDVVVAADPRRC 480
QY 481 LGLLVPRLLDRCALSGLKEASDAEVLASEPVRAMPADMLKRLNREATGNASRIIMVGL 540
DB 481 LGLLVPRLLDRCALSGLKEASDAEVLASEPVRAMPADMLKRLNREATGNASRIIMVGL 540
QY 541 LDTPPSIDKGEVTDKGSINQRAVLQWRSKAVDALYRGEDQSMLRDEATL 589
DB 541 LDTPPSIDKGEVTDKGSINQRAVLQWRSKAVDALYRGEDQSMLRDEATL 589

RESULT 2
US-09-750-986D-30
; Sequence 30, Application US/09750986D
; Publication No. US20030228670A1
; GENERAL INFORMATION:
; APPLICANT: Steinbuechel, Alexander
; APPLICANT: Pfelefer, Horst
; TITLE OF INVENTION: SYNTHETIC ENZYMES FOR THE PRODUCTION OF CONIFERYL
; TITLE OF INVENTION: ALCOHOL, CONIFERYLALDEHYDE, FERULIC ACID, VANILLIN AND
; FILE REFERENCE: Bayer-9998-CAO
; CURRENT APPLICATION NUMBER: US/09/750, 986D
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: 196 49 655.1 GERMANY
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 30
```

LENGTH: 589
; TYPE: PRT
; ORGANISM: Pseudomonas sp.
US-09-750-986D-30

Query Match 100.0%; Score 3082; DB 10; Length 589;
Best Local Similarity 100.0%; Pred. No. 7.7e-289;
Matches 589; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MRSLEALPPPGILRLERLHMAKTRPEQTCVAAARAANGEMRRISYAEFHNVAIAQSLL 60
DB 1 MRSLEALPPPGILRLERLHMAKTRPEQTCVAAARAANGEMRRISYAEFHNVAIAQSLL 60
QY 61 PYGLSARPLLIVSGNDLEHLQLAFGAMYAGIPYCVSPAYSLSDLAKLRHIVGLLP 120
DB 61 PYGLSARPLLIVSGNDLEHLQLAFGAMYAGIPYCVSPAYSLSDLAKLRHIVGLLP 120
QY 121 GLVFAADAAFPORAETIIPDDVPALFTRGELAGRTVSFDSLLEPGGIEADNAFAATG 180
DB 121 GLVFAADAAFPORAETIIPDDVPALFTRGELAGRTVSFDSLLEPGGIEADNAFAATG 180
QY 181 PDTIAKFLFTSGSTKLPAKAVPTTQRMLCANQOMLQTFPVFGSEPPVLDMLPMNHTFGG 240
DB 181 PDTIAKFLFTSGSTKLPAKAVPTTQRMLCANQOMLQTFPVFGSEPPVLDMLPMNHTFGG 240
QY 241 SHNIGIVLVNGGTYLDDGKPTAQGFATLRNLSEISPTAYLTVPKGMEELVGALERDST 300
DB 241 SHNIGIVLVNGGTYLDDGKPTAQGFATLRNLSEISPTAYLTVPKGMEELVGALERDST 300
QY 301 LRERFPAARMKLFPPAAAGISOGIMDRLDRAVEOHCGERIRMMAGLQMTETAPSCFTTGP 360
DB 301 LRERFPAARMKLFPPAAAGISOGIMDRLDRAVEOHCGERIRMMAGLQMTETAPSCFTTGP 360
QY 361 LSMAGYIGLPAPCEVTLVVDGKLEGRFHGPHVMSGYRPAEONQAQDEEGYCSGDA 420
DB 361 LSMAGYIGLPAPCEVTLVVDGKLEGRFHGPHVMSGYRPAEONQAQDEEGYCSGDA 420
QY 421 IKLADPADPQKGMFDGRIAEDEKLSGVSFVSGPLRTAIVLEGGSYLDVVVAADPRRC 480
DB 421 IKLADPADPQKGMFDGRIAEDEKLSGVSFVSGPLRTAIVLEGGSYLDVVVAADPRRC 480
QY 481 LGLLVPRLLDRCALSGLKEASDAEVLASEPVRAMPADMLKRLNREATGNASRIIMVGL 540
DB 481 LGLLVPRLLDRCALSGLKEASDAEVLASEPVRAMPADMLKRLNREATGNASRIIMVGL 540
QY 541 LDTPPSIDKGEVTDKGSINQRAVLQWRSKAVDALYRGEDQSMLRDEATL 589
DB 541 LDTPPSIDKGEVTDKGSINQRAVLQWRSKAVDALYRGEDQSMLRDEATL 589

RESULT 3
US-10-369-493-20490
; Sequence 20490, Application US/10369493
; Publication No. US2003023675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 20490
; LENGTH: 575
; TYPE: PRT
; ORGANISM: Rhodospseudomonas palustris
US-10-369-493-20490
```

Query Match	47.3%	Score 1457;	DB 14;	Length 575;
Best Local Similarity	50.5%;	Pred. No. 1.5e-131;		
Matches 292;	Conservative 86;	Mismatches 194;	Indels 6;	Gaps 2;

Qy	1	MRLEALPPGGIILRLREHMAKTREQICVAARANGCWRIRISYEMPHNRALAAQSLI	60
Dh	1	VRSTTTLTYDPAVITIDRLNHFAETADRDVPMERNCEGMRIRIAYEMLRRAQTLASALI	60
Qy	61	PYGLSARPELLIVSGNDLEHLQAFGAMYAGIYPCVSPAYSLSLQDLAKLRHIVGLLP	120
Dh	61	ARGLSARPMVILTSGNSIDHAMVFGALYAGVMCEVSPSYISVSKDYGLKHIVGLLTP	120
Qy	121	GLVFAADAAFPQAIETILPDDVPALFTTGELAGRTVSPDSLBPQGITEDNAPATG	180
Dh	121	GLVFAADITAFABAIATATVPEDVELATRGVEVGRKVTSLABELLATPEHPELAARHAI	180
Qy	181	POTIAKPLFTSGSGTKLPKAVPTTORMICANQOQLTTPVFGGEPPVLVDMLPMNHTFG	240
Dh	181	HDTIAKPLTSGSGTGNPKAVINTORMICANQVMIRBMAFLDEDPVYIDMLPMNHTFG	240
Qy	241	SHNIGIVLYNGCTYYLDDGKPTAQGEAETLRNLSEISPTAYLTPVKGWELVALERDST	300
Dh	241	NHNIGLTLFNGSGMYIDDGKPTAGIASTIRNLREIATPYEVNVPKGYESLLPVLREDQ	300
Qy	301	LRERFFARMKLPFPAAAGLSQGIWBLDRVYAEQHCERIRPMAGLGMTTASCTFTTGP	360
Dh	301	LRKLFFSRLLAMFFSGASLAAHVNGIDEVAAVETARVPMITGLCATETAEFFMVSVPQ	360
Qy	361	LSMAGYIGLPAPCEYKVLVPVDGKLEGRFHGPHVMSGYVRAPQUNOADEEGYCSGDA	420
Dh	361	TSSRGHGLVPPNENKLVPPNNGKLEVRKAGPVITGGVRRAPBELTDKADDEBGFYKXDA	420
Qy	421	IKRLADPADPQKGLMFDGRILAEDFKLSGVFVSGVPLRTAVLVEGSGYYLDDVVAAPDREC	480
Dh	421	LKRVADANDLSRGDPFQGRISIEDFKLASGTMVSVGPLRAKFLAACSLVDVVIAGLDROY	480
Qy	481	LGILLVPRLLD---CRALSGLKEASDAEVLASBPVRAMPADMKRLNREATGMSRIYW	537
Dh	481	VTALA---IIDDPGCKILINATLPLEDLAGMAADHLIREAFRRERFATLTLQATGSSNRVTR	537
Qy	538	VGLIDTPPSITDKEVYTDKGSINORAVYLQWRSAYVDLY	575
Dh	538	AVLLGPEPLSIDKBEITDKGSVNOQRAVLEVPASILADLY	575

```

1      RESULT 4
2      US-10-369-493-11697
3      ; Sequence 11697, Application US/10369493
4      ; Publication No. US20030233675A1
5      ; GENERAL INFORMATION:
6      ; APPLICANT: Cao, Yongwei
7      ; APPLICANT: Hinkle, Gregory J.
8      ; APPLICANT: Slater, Steven C.
9      ; APPLICANT: Goldman, Barry S.
10     ; APPLICANT: Chen, Xianfeng
11     ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
12     ; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
13     ; FILE REFERENCE: 38-10(52052)B
14     ; CURRENT APPLICATION NUMBER: US/10/369,493
15     ; CURRENT FILING DATE: 2003-02-28
16     ; PRIOR APPLICATION NUMBER: US 60/360,039
17     ; PRIOR FILING DATE: 2002-02-21
18     ; NUMBER OF SEQ. ID NOS: 47374
19     ; SEQ. ID NO 11697
20     ; LENGTH: 576
21     ; TYPE: PRT
22     ; ORGANISM: Agrobacterium tumefaciens
23     ; US-10-369-493-11697

```

Query Match	42.0%;	Score 1294;	DB 14;	Length 576;
Best Local Similarity	46.5%;	Pred. No. 9e-116;		
Matches 266;	Conservative 84;	Mismatches 220;	Indels 2;	Gaps 2

Qy	5	EALLPFGRLLELEHMAKTREQOTVAAARAANGERRRISYEMHNPALIQSLPGL	64
Db	1	DPGPTPQKNEILLHMCRSABERTMMADROGREPRRRSVAYEALDKIRRIQGLFLDDHL	60
Qy	65	SAERPLLIVSGNDLEHLQLAFGAMTAGIDPYCEVSPAYSLISODLAKLRHTVGLLOPGLVF	124
Db	61	SVBRPLLVISENSIIBALMWLAQHVGISAATPAVA-TSADLTKLDIRQITPGWVF	119
Qy	125	AADAAAPQRALEITLLEDDVPAITFTREBLAGR-RITVSFDSLLEQPGIEADNAFAATGSDT	183
Db	120	AEATPERRALGVEFDDGPTPLVGLRLPEBRNANTHIFETLLETPTEAVDRAFDVAGPDT	179
Qy	184	IAEFLFSSGSTKLPKAVPTTQRLCMNQOMLQTEPFVGESEBPVLVDMLEPMHTEGCSHN	243
Db	180	YAKFLFTSGTSGPKAVIQTQRLCSNOEMIDCICYFPEEBPPVYVDNAPMHHTAAGKV	239
Qy	244	IGIVLYNGTGYLLDDKPTAOGFAETLRNLSEISPTAYLTVPKGMBELVGALEDRSTLRE	303
Db	240	FNLYLVNGTGYTYIDRKPSPAQIGQLTDLNRDISPTWYFNVPAGHEMLVQAMRKDEALCR	299
Qy	304	RFRPRKMLFFFAAAGLSOGIMWRDLRVAQHCHGERRMMAAGLGMETAPASCFTTGPPLSM	363
Db	300	SFRDRLKMLMYAGAGAAOHTWDALTELSPATGHAULMAGAGISTETAFSLFCTEPDOK	359
Qy	364	AGYIGLPAPGCEVKLPVVDGKLEGRFHGSHNMGVYARBEQNAQOFDEEGYVCSGDATKL	423
Db	360	PGNIGTPAOQVTKLVPFDGRYTELRKGNITIPGWYRNGELTAAAPFDEGFRIGDYKFE	419
Qy	424	ADPADPOKGLMPDQIAEDPFLKSSGVFVSGLPRTAVALDEGGSYYLDDVVAAAPDEECGL	483
Db	420	AYADDDPRRGYTFPGMAENPKUQTGTWVAVGLPRAQLVMPFAGLLRDVAITTEGNBAELGA	479
Qy	484	LVPRLRLDCRALSGLGKSAASDAEVLASEPVYRAWFADMLKLRERATGNAISRIMWGLDPT	543
Db	480	LVPFIPALRELYRGSOHLSDAEIITHPVSRAQIVAKLSAHOXKQASGSASRYMILYMED	539
Qy	544	PPSIDGCEVTDKGSINORAVLIQWRSAKCYDALY	575
Db	540	ALAPKEVEVTDKGSINORAVLIHKEKELVSLY 571	

```

RESULT 5
US-10-369-493-14740
: Sequence 14740, Application US/10369493
: Publication NO. US20030233675A1
: GENERAL INFORMATION:
: APPLICANT: Cao, Yongwei
: APPLICANT: Hinkle, Gregory J.
: APPLICANT: Slater, Steven C.
: APPLICANT: Goldman, Barry S.
: APPLICANT: Chen, Xi'anfeng
: TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
: TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
: FILE REFERENCE: 38-10(52052)B
: CURRENT APPLICATION NUMBER: US/10/369,493
: PRIOR FILING DATE: 2003-02-28
: PRIOR APPLICATION NUMBER: US 60/360,039
: PRIOR FILING DATE: 2002-02-21
: NUMBER OF SEQ ID NOS: 47374
: SEQ ID NO 14740
: LENGTH: 576
: TYPE: PRT
: ORGANISM: Agrobacterium tumefaciens
US-10-369-493-14740

```

Query	March	Score	1294	DB	14	Length	576
Best Local	Similarity	46.5%	Pred.	No	9e-116		
Matches	266	Conservative	84	Mismatches	220	Indels	2
Gaps							2
QY	5	EALLPFRILIERLEHMAKTRPEQTCVAAAPAAANEWRISIAEYFHNVRATASLIPGL	64				
DB	1	DLPLGPPKDLNRLHHCGRSAPEKRTIMADRCGRPEWRVSAEALDKTKRIGQFLDIDL	60				

Qy	65	SARRPLIVSGNLEHQLAFGMVYAGIRPCVSPRYSLSDQIATKRIVGLQGLV	124
Db	61	SVERPLIVSENSIEHMLVLAQHVGISAAITPVA-TSADLTTLADI RQIITGWGF	119
Qy	125	AADAAPFORAIFETLPDDVPALFTREBELAGR-RTVPFDBLLEOPGIEADNAPAAATGPD	183
Db	120	AEDATPFRALAGEVFDGTPVLVGLRMLPEDRSMTTFHEITLLETPTEAVADRAFVAGPDT	179
Qy	184	IATETLFSSGSKPKPAVPTTQRMLCANQOMLOTFFVGEPEEPVLYDMLPMNHTGGSSHN	243
Db	180	VAKFLFTSGTTGSPKAVIQTORMLCSNOEMIADCYFEEBEPVVVDMAPMNHTAAGKV	239
Qy	244	IGIVLYNGGTYVDDGFEPTAQGFATLRLNLSIEISPAVYTVPKGMEELGALERDSTLRE	303
Db	240	FNLVLVNGGTYIIDRGKSPAQIGQTLDNLRDISPTWYFNVPAGHEMLVQAKRDBALCR	299
Qy	304	RFPARKMLFFPAAAGLSQGIWDLRYVAEOHCSEIRIMNAGIGMETAPASCFTTGPISLM	363
Db	300	SFPEDLMLMYAAGMAQHTWDLTBLISMAVYSHAALVMAAGISGETAPSLFCTEPODK	359
Qy	364	AGYIGLPAPCEVYLVPVDDGKLEGRFHGPHVSVGVRAPEDQNAOAFDEEGVYCSGDATKL	423
Db	360	PGNIGIPAGQVMTKVLVPFDGRYELRLKGNITPTGVYRNBELTAAAFDEEGFRIIDITYKF	419
Qy	424	ADPADPOKGLMPFGRIADDFKLSGGYFVSVYGLRTPAVLBEGGSYVLDDVVAAPDRECLG	483
Db	420	AVADDDPRGEGYFGGRMAENFKLQGTGVAVAGPLRAOLVMMFAGLIRDAITYGENRAELGA	479
Qy	484	LVFPRLLDCALSGLGKEASDAEVLASEPRAWFADMLKRLUREATGNASRTIMWGLDPT	543
Db	480	LVEFPIFALRELYVRGSHLSDAEIIHHSVRAQIVAKLSAHQKQSGSASRWIRLIVWED	539
Qy	544	PPSIDKGEVTDKGSINQRAVLTQWRSKAVDALY	575
Db	540	ALRFEKEGVTDKGSINQRAVLTIRKEIVLSY	571

```

RESULT 6
US-10-369-493--15216
Sequence 15216, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ. ID NOS: 47374
SEQ ID NO 15216
LENGTH: 576
TYPE: PRT
ORGANISM: Agrobacterium tumefaciens
US-10-369-493--15216

```

Query Match	42.0%	Score 1294	DB 14	Length 576
Best Local Similarity	46.5%	Pred. No. 9e-116		
Matches 266	Conservative 84	Mismatches 220	Indels 2	Gaps 2

QY	125	AADANPFOALFETILIPDDVPAFTFGELAGR-RTYSFDSLLAQPGCIEADMNFAATGPDT	183
Dd	120	AEDAPFPFRALGGEVDDGTPLVLGNLNPEDRSNTFFETLLETPTEAVDRAADAAGPT	179
QY	184	IAPFLFTSGSTLCPXAVPPTTCOMLCANOQMILLQTFPEVFCGEESPVLVDWLPWNHTFGGSHN	243
		: :	
Dd	180	VAKFLFTSGTTGSFXAVIQTOTMCLSNQEMIADCYGFPREPPVVWDMAPMHNTAAGNKV	229
QY	244	IGIVLYNGCTYYLDDGKRPACGSAFTLRNLSEISPAIYLTVPKGMELYGALERDISTIRE	303
		: :	
Dd	240	FNLVLYNGTYYIIDGKSPQAIOGTLDLNRDISPWYNFNVAAGHMLVOAMKCDALCR	299
QY	304	RFFAAKCFPFAPAAGLSOGIMDLRNVAAEQHCGERIRMMAGCMETETAESCPTTGPLSM	363
		: :	
Dd	300	SFFRLKLMTLVAGAAGMOITMDALIELSMATYGHAVLMAGLGISTTIAFPSLFCTERPODK	359
QY	364	AGYIGLPAAPGCEVKLVPVDDGLEGRFHCBPHVMSGYWRABEQNAQAFDEBGYYCSSDAIKL	423
		: :	
Dd	360	PCNIGIPAOGVYMKLVPPDGRYELRKGPNIIFPYWRNDELTAAPADEEGFYRICBTVPK	419
QY	424	ADPADPOGLMPDGIADBFKLSGVFVSVGELRRPAVLLEGSSYVLVDVVYAANDREBCLG	483
		: :	
Dd	420	AVADDPRRGFYFDGWAEINFUKQTGTWAAVGLPQLVMMFAGLIRDAYITGENRELGA	479
QY	484	LVPFRLDCRALSGHKEASDAEVLASBEVRAMPDWLKRINRELATGNARIMVGLDIT	543
		: :	
Dd	480	LVPFRTPALRELIVRSSQHLSDAEILIRHPSVRAQIYAKKSAHQKAGSASRVRIIMVED	539
QY	544	PPSIDKGEXTDKGSINORAVLDMBSAKDYALT	575
		: :	
Dd	540	ALRFEPKEGVTDKGSINORAVLLHRHELEVESLT	571

```

RESULT 7
US-10-369-493-14320
/ Sequence 14320, Application US/10369493
/ Publication No. US20030233675A1
/ GENERAL INFORMATION:
/ APPLICANT: Cao, Yongwei
/ APPLICANT: Hinkle, Gregory J.
/ APPLICANT: Slater, Steven C.
/ APPLICANT: Goldman, Barry S.
/ APPLICANT: Chen, Xianfeng
/ TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
/ TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
/ FILE REFERENCE: 38-10(52052)B
/ CURRENT APPLICATION NUMBER: US/10/369,493
/ CURRENT FILING DATE: 2003-02-28
/ PRIOR APPLICATION NUMBER: US 60/360,039
/ PRIOR FILING DATE: 2002-02-21
/ NUMBER OF SEQ. ID NOS: 47374
/ SEQ. ID NO 14320
/
/ LENGTH: 553
/
/ TYPE: PRT
/
/ ORGANISM: Agrobacterium tumefaciens
US-10-369-493-14320

```

Query Match	40.6%	Score 1250;	DB 14;	Length 553;
Best Local Similarly	46.2%	Pred. No. 1.5e-11;		
Matches 256; Conservative	82;	Mismatches 214;	Indels 2;	Gaps 2;

QY 5 EALLPFGGILRLERHMAKTRPEOTCVARAANEMRRISYAEMFHNVAIAOSLLPYGL 64
DB 1 DPLGPPQKILNERLLHWCSSAPEKTTWMDRQGRPPRRVRSYAELDKIRIGQFLDHL 60

Qy	65	SAEPRILVSNDELEHQLQARFAMAYAGI	PCYCPBPAYSL	SODLAKXHIYGLQPGVLF	124
		: : : : :	: : : : :	: : : : :	
Db	61	SVERLELVLSNLSIEHLMVTAQGVGASALIPPA	-TSADLT	KLADIGQITPGWVF	119
		: : : : :	: : : : :	: : : : :	
Qy	125	AADAPFORALETLLPDDVPAL	FTFRGELAGR	-RTVSPDLSLEQPGCIEADNAPAAATGPD	183
		: : : : :	: : : : :	: : : : :	
Db	120	AEDTPTFFRRALAGEVADDOCTPLVGR	RNNLPEDSNFHFPEFTLLETERTEA	VDASAFDVA	179
		: : : : :	: : : : :	: : : : :	

	Query Match	31.5%;	Score 972;	DB 14;	Length 621;
	Best Local Similarity	38.3%;	Pred. No. 1,5e-84;		
	Matches	219;	Conservative	86;	Mismatches 249;
				Indels	18;
				Gaps	6;
Qy	11	PGRIIERLEHMAKTRPEQTCVAAARAAN-GEWRISVAMEFHNNVAINQSLIPYGLSAREP	69		
	50	PRSIAMHLIAKAAAEHEDRRPYLKKOREBNHCGWRGCTYEABRAHVAEGVAAQMLDQKIQPHDS	109		
Qy	70	LLIVAGNDIEHQLQAFGAMYAGIPYCPVSPAYSLISQDLKLRHIVGLQGLVFPAADAA	129		
	110	VMILSGNSIEHMLMTLGATYAGVPAAPISPAVSLISIDHAKLKRCEFEKVARVVFPAOSGA	169		
Qy	130	PFQRAIETLLPDDVPAIFTRGELAGRTVSEFDSLLEOPGIEADNAEPAATGPDITAKLEF	189		
	170	MFAGALATILKALDPSLIVITADAGAGSALITFAEVAATTPPAAVAAADAVAPATVAKLFL	229		
Qy	190	TSGSTKLPAVAVPTTQMLC--ANQOMLLOTFPVYGEPEPVYLVWLPMNHTFGGSNIGI	246		
Db	230	TSGSTGLPAVAVQTHQMMAGVLIAGQSLRTDVPF--DEVQSTLEMMWPSHISAQNIENGA	287		

Query Match	12.4%	Score 383.5;	DB 14;	Length 544;
Best Local Similarity	25.0%;	Pred. No. 1,1e-27;		
Matches 142;	Conservative 108;	Mismatch 232;	Indels 87;	Gaps 23;
QY	41	RRISYAMFNNVNAIQSSLPYGLSABRPLLIYSGNDLEHLQAFAMVAGIYPCVSPA	100	
DB	28	RQLSWROLCSIAIRIYMHENLVNTGQRVVTISHNCMBLAEFCQIKGAIYAV---	84	
QY	101	YSLISQDLAKIRHIVGILQGLVFA-----ADAAPF-----QRA	134	
DB	85	YANSEEM--MYTCLSLAEPALVITEKSLAQTLAAGSPWQWVLTQAGPFDPFGIDIE	142	
QY	135	IEITLLPDDVAIFTRGELAGRTVSPDSLQEPGIEADNAFAATGPTITAKFLFTSGST	194	
DB	143	TDLISPDKAVSHEAIGF-ALLRAITTDLQATQ-----AEQAF-----LMFTSGSS	187	
QY	195	KLPAVPTTQRMCLCANDQMLQTFPRVPGSEBPVLVDMLEPMTFFGGSNIGIVLVNGTV	254	
DB	188	GKPGGVLLSSQNNI--LSQQALQTLIMHF--DGQSFSLTELPWHHSFGCLFEKYTAIFSQAML	245	
QY	255	YLDGGRKTAGFAETLRNLSGISPTAVYLVRKCMESLVGALBERDSTLRERPFARMLPF	314	
DB	246	HIDNR--GMDIDRMCTNSQLKPTFPSPKILIVVANKMKGSEAFHRAVAPTLQMTS	303	

QY		315	AAAGISGQIMRLDRAVBOHGERIRMAAGLMTETAPRCTTTTGLSMAGY-----IGL	369
Db		304	AASKLPPIEDYF-----KHCG--IYAEGMGITESTP-CLTIKGP----GYEHNVSGR	351
QY		370	PAPGCEVLKVPVDGKLKGFRPHMYSGYRAPAEONAOAFDEBGYYCSGDATKIADPADP	429
Db		352	PLPNYLILSDSGSEILA--CPNWLGYADEAANQRCF-RQGERTTDLGKMI-----	403
QY		430	OKG-LMPFGRIADPEKLSGSVFVSGPLETRAVLEGGSVLVDDVVAAPRECCGLLVF--	486
Db		404	-KGELIVLLIGRLDSVKYLSNGEKVSSDAIE-QSLFEKTDIINFAIIDIEOPRYATALLFIN	461
QY		487	PLLDCRALSGIKGEASDAEVLAASEVBRAFMDKRLRLREATGNASRIMTVGLIDTPPS	546
Db		462	PEL--CRGRFG-----ENCTLGHSRLGHIAIAVVQQHQVEDSSFKLIAVAVTETSIS	514
QY		547	IDKGVPTDKSGINQRAYLQMRSAKYDALY	575
Db		515	LKRGVTPSPFKVSQRTVRNNYQADVADALY	543

```

RESULT 10
US-10-156-761-13094
; Sequence 13094, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156, 761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 13094
; LENGTH: 636
; TYPE: FRT
; ORGANISM: Streptomyces avermitilis
; JS-10-156-761-13094

```

Query Match	10.8%;	Score 333;	DB 14;	Length 636;
Best Local Similarity	23.7%;	Pred. No. 1.1e-22;		
Matches 151;	Conservative 100;	Mismatches 261;	Indels 124;	Gaps 25

```

QY      26 PEOTVAAARAANGEMRRISVYAMFNHVRVIAQSILPFGYSAEPPILVSGNDLHEHQLAF 85
QY      45 PEQGAAPAGQCPDDMKLSLSMAQAAERVPVAAAGLITLGYQPEQRVALASTRYEMILLAD 104
Db      86 GAMTAGIPYCPVSPAYSLISODLAKRHLVGLLPQGLVPADAPFORALE----- 136
QY      105 GIMCAGAAATTITYP-----QTNADSAFILLADSESKVLLIAEDPAQAKAVEKREAPALTL 159
Db      137 ---TLIPDVPA---IFTRGELAGRTVSF-----LSLLPQGGIEADNMFAATGPPITIK 186
QY      160 HAVVVIDPFGVESSDWVLTIAELERAGAAHLKKNPDILKEKVGATSDQ-----LAT 210
Db      187 FLFTSGSTRKLPKAVPTTQRMLCANQOMLQTFEPVFG--EERPVLYVMLPWNHTFGGSHNI 244
QY      211 LIYSGTIGRPRGCV---RLPHDNMSYMAKATAGTGLIADDDQVLYMLPLAHVFGK---- 262
Db      245 GIVLVNGGTYIYDDGKPTAQC--FAETLANLSISTAYLVTPKGMELVGALERDSTLR 302
QY      263 --VLISG---QLEVGHVTAADGRVVKIENLPVQOPTYMAAVRIRIEKXNGVA--AKAR 315

```

[illegible]

```

RESULT 11
US-10-369-493-10386
; Sequence 10386, Application US/10369493
; Publication NO. US2003023675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 36-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 10386
; LENGTH: 582
; TYPE: PRT
; ORGANISM: Cytophaga hutchinsonii
; US-10-369-493-10386

```

Query Match	10.5%;	Score 324;	DB 14;	Length 582;
Best Local Similarity	23.6%;	Pred. No. 7.2e-22;		
Matches 144;	Conservative 101;	Mismatches 264;	Indels 100;	Gaps 26

```

Oy      26 PEOTCAARAANGEMRIRIYVAMFHNVAIAQSLIPYLSAERPLIYSGNDLPHLOAF 85
Db      13 PKDQALATK-ENGWTWISYSYVQELIDKVSIGFKSLQDQDKVAIISPRRPMNIDL 71

Oy      86 GANYAGI PYCPVSPAYSLISODLAKRHIVGLQPLVPAADAAPFORAETIIPDDV-P 144
Db      72 GLQOIAVAVSPTIYP--TLTVE--YRYIFDDAEKVMFVVDQOLFAKATAVNGSRNP 126

Oy      145 AITRRELIGRRIVSFDSLLE--OPGGAEADNAFPAATPDIIAKFLPFGSGTKPKRVPT 202
Db      127 VITLIDQVNGASKWDITIALDQGPASW-LDPYKRVANNDULLIYISGTTGPKGVM 185

Oy      203 TORMLCANOQMLOTPEPVGEEPPVLVDMLPNWHTFGGSHNIGIYLY--NGGTYYLDGK 260
Db      186 THNNIISNVAIAPIIMPV--NNNHKALSPLEPCHY--ERMCLCYLYLSQGVSIYAE 240

Oy      261 PTYQGAETILIRUNSEISPTAYIITVKGME-----ELVG-----ALER 297
Db      241 DTVG-----ENIKEQPMFVITVRLLEKVDKIIVAKGALGTGKFLFYMALEIGKY 294

```


KM	Sequence:	632 AA;	68280 MW;	C082A111469934F7 CRC64;
QY	Query Match	47.3%;	Score 1457;	DB 2; Length 632;
Db	Best Local Similarity	50.5%;	Pred. No. 1.1e-97;	
	Matches 292;	Conservative 86;	Mismatches 194;	Indels 6; Gaps 2;
QY	1 MRSLEALLPFGRIILERLEHNAKTRPEGTCVAAABAANEMBRISVAEMFNHVRAIQSLI	60		
Db	46 VRSITTLTDYVRITRDLHFETAPDPVFMERERGEGMRIRISAEMLRAQAQTASLTI	105		
QY	61 PYGSAERPLLIIVSGNDLEHLQAFGANAYAGIPYCVPSPAYSLSODLAKRHIVGLDLP	120		
Db	106 ARGSAERPVMILSGNSIDHAMVMRGALYAGVAMCPVSPPSLVSKDYGKLRIHIGLLTP	165		
QY	121 GLVPAADAPAPQRAIETLLPDVDPAIFPRGELAGRRTYSFDSLSLEGPGIEDNAPFAATG	180		
Db	166 GLTAPADDTTAAPALATLATVPEDVELAARGEVGKKVTLSLAELLATPEHPPELLAKGAIG	225		
QY	181 PDTIAKFLFTSGSTYLPAKAVPTTQRMLCANDOMLTQTFPVFGEEPVLVDMLPMNHTEGG	240		
Db	226 HDTIKAFLLTSGSTGNPKAVINTQGMCANQMIREMAAPLKDEPPVTVDMIPMHHTFGS	285		
QY	241 SHNIQIVLYNGCTYYLDGDKPTAGCGFAETLRNLSEISTPAYITVPKGMEBLVALDERST	300		
Db	286 NHNIQGLTFENGGSMTVIDDGKPTPAGIASTINRLREIATPVFFNVPKGYESLPLVLRDDQ	345		
QY	301 LREPFPAAMKLPFPFAAGLSGIMDRUDRVABQHCGEIRIMMAGIMETA PSTCFTTGP	360		
Db	346 LRKLPFFSKHAMFFSGASLAHWMMGLEVAALGTGAIPMLTGATIELTDFMSVTPO	405		
QY	361 LSMAGYIGLPAPGCEVKLVPPVDGKEGRFHGHVWMSGWRAPEDQNAOFDEBGYYCSGDA	420		
Db	406 TSRGSHVGLPVPGNEAKLVPPNNKGLEVAKGPNTIPGWRAPELTDKA FDEBGFYKLANDA	465		
QY	421 IKLADPADPQKGLMPDGRHIAEDPKISSGVPPSVGPIKTRAVLEGSSYLDVVYAAPRDEC	480		
Db	466 LKPVDANDLSRGFDGRISEDFKIASGTWVSVGLRKAKFLAACSLVRDVVIAQLDIDY	525		
QY	481 LGILVFPEPLD---CRALSIGKEASDSEVLASEBVRAFMDLRLNREATGNASRIWM	537		
Db	526 VTALA---ILUPDQGLKNATLPLEDLGMAADHIIRAFNERPATLLTQATGSSNRVTR	582		
QY	538 VGILDTPPSINDKEVTDKGSINQRAVLQWSAKVDALY	575		
Db	583 AVLLGEPLSIDKGEITTDKGSVQVRVLEYRASLIADLY	620		
RESULT 13				
O89CF2	PRELIMINARY;	PRT;	627 AA.	
AC	O89CF2;			
DT	01-JUN-2003 (TREMBLrel. 24, Created)			
DT	01-JUN-2003 (TREMBLrel. 24, Last sequence update)			
DT	-01-MAR-2004 (TREMBLrel. 26, Last annotation update)			
DE	Long-chain-fatty-acid-CoA-ligase.			
EN	OrderedLocusNames=blt7845;			
OC	Bradyrhizobium japonicum.			
OC	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;			
OC	Bradyrhizobiaceae; Bradyrhizobium.			
OX	NCBI_TaxId=375;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=USD110;			
RX	MEDLINE=2248498; PubMed=12597275;			
RA	Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiuni T.,			
RA	Sasamoto S., Watanabe A., Idegawa K., Iitguichi H., Kawashima K.,			
RA	Kohara M., Matsuno M., Shimpō S., Tsurutoka H., Wada T., Yamada M.,			
RA	Tabata S.;			
RT	"Complete genomic sequence of nitrogen-fixing symbiotic bacterium			
RT	Bradyrhizobium japonicum USD110."			
RL	DNA Reel. 9:169-197(2002).			
CC	-1- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme			

CC	family.	EMBL: AP005963; BAC53110.1; -
DR	GO; GO:0016874; F:ligase activity; IEA.	
DR	GO; GO:0008152; F:metabolism; IEA.	
DR	InterPro; IPR000873; AMP-bind.	
DR	Pfam; PF00501; AMP-binding; 1.	
DR	PRINTS; PR00154; AMPBINDING.	
DR	PROSITE; PS00455; AMP BINDING; 1.	
KW	Complete proteome; Ligase	
SC	SEQUENCE 627 AA; 68213 MW; F8C8C8C304D438F88 CRC64;	
Query Match	45.5%; Score 1401; DB 2; Length 627;	
Best Local Similarity	50.7%; Pred. No. 1,3e-93;	
Matches	293; Conservative 80; Mismatches 199; Indels 6; Gaps 4;	
QY	1 MRSIEALLPFGCRILRELEHNAKTRPEQTCVAAAPAAENEMRISYAEKFNHVRATASGL 60	
DB	41 LRPQGLADGVSVRTDRLHMWATTTPDVFPAEREGGGRMKRTYAEILLTSRHSIALSLI 100	
QY	61 PYGSAERPELLIVSGNDLEHLQLAFGAMYAGIPIPCVSPASLSLSDLAIRHIVLQAP 120	
DB	101 ORGISADDPVILISGNSIDHALLAFGAFAVAPCPVSPASLSVSKOVGKLSYLMKLLTP 160	
QY	121 GLVPAADAPAPQRAI--ETIIPDDVPAIFTRGELAGRTTSPDSLDEPGGIEADNAFA 178	
DB	161 GLVPAEADADKPSDLAANVSLGTETIASY--GHVAGROVTLADIMAPIRGDIDEVHGK 218	
QY	179 TGPPTIAKFLFTSGSTKLPKAVPTTORMLCANQOMLOTPEVFGSEEPVLVDMLPMNHTF 238	
DB	219 IGPDTIAKFLFTSGSTGNPKAVITORMLCANQOMLRETLAFKDEPVIITDMLPMNHTF 278	
QY	239 GGSNHNIGIVLVNCGTYLLDDKPTAOCFAETLRNLSEISPTAYLITVPKWEELVGLERD 298	
DB	279 GGNHNIGITLVNCGSMYLDACKPMPGGIEETVRNLQISPIVYFVNPKGYSLLPEYLRDD 338	
QY	299 STAEERPAARKLPPFAAAGISQGIWRLDVAQAQHCERIRMAAGCMTETAPSCITTF 358	
DB	339 QGLAKKFFDRLHAMFFSGAALSPVPMOSLDLAVYEKRYRVPMLGLGATETAPFPMNVN 398	
QY	359 GPLSMAGYIGLPARGCEVYKLPVVGKLEGRFHGPHVMSGVRAPEONAAQFDEEGYCSG 418	
DB	399 PRTSRSGHVGLPVGNDAKLVPNNGKLEVRAKGPNVMPGVWRQADITAKSPDEBGFYKLG 458	
QY	419 DATKADPADPQKGLMPDGRJAEDPKLSSGVFVSGPLRTAVALLEGSSYVLDVVAADDR 478	
DB	459 DALKPADPDDLNAAFDEPDGVRVSEDFKASGFWSVGPRARAFVAAACLVADVDVYAGINR 518	
QY	479 ECLGLVPEPRLLDGRALSGLGKEASDAEVLASEP--VRAMPDWMLKRLREATGNASRIIM 537	
DB	519 DEVALAVLVLDIDGCRLVNPT--LPADDLVVAARDRLVRAFAFERLTRLFGSATGSSTRITR 577	
QY	538 VGLLDTPEPSIDKGEVTDKGSINQRAVLQMSAKYDALY 575	
DB	578 AILMDTPLSIDKGEVTDKGSINQRAVLLEHNGALIDELY 615	
RESULT 14		
Q7CZA8	PRELIMINARY; PRT; 591 AA.	
ID	Q7CZA8	
AC	Q7CZA8;	
DT	05-JUL-2004 (TREMBlrel. 27, Created)	
DT	05-JUL-2004 (TREMBlrel. 27, Last sequence update)	
DT	05-JUL-2004 (TREMBlrel. 27, Last annotation update)	
DE	AGR_C_2614P.	
GN	OrderedLocusNames=AGR_C_2614;	
OS	Agrobacterium tumefaciens (strain C58 / ATCC 33970).	
OC	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;	
OC	Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.	
OX	NCBI_TaxId=176299;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=Cereon;	
RX	MEDLINE=21608551; PubMed=11743194;	

[illegible]

```

RESULT 11
Q6N8W6 ID Q6N8W6 PRELIMINARY; PRT; 632 AA.
AC Q6N8W6;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Putative feruloyl-CoA synthetase (EC 6.2.1.-).
GN Name=fc82; OrderedLocustNames=RPA1787;
OS Rhodopseudomonas palustris.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Rhodopseudomonas.
OX NCBI_TaxID=1076;
RN [1]
RP
RP SEQUENCE FROM N.A.
RC STRAIN=CGA009 / ATCC BAA-98;
RC Pubmed=14704707; DOI=10.1038/nbt923;
RA Laimer F.W., Chain P., Hauser L., Lamerdin J.E., Malfatti S., Do L.,
RA Land M.L., Pelletier D.A., Beatty J.T., Lang A.S., Tablita F.R.,
RA Gibson J.L., Hanson T.E., Bobst C., Torres y Torres J.L., Pers C.,
RA Harrison F.H., Gibson J., Harwood C.S.;
RT "Complete genome sequence of the metabolically versatile
RT photosynthetic bacterium Rhodopseudomonas palustris";
RL Nat. Biotechnol. 22:55-61(2004) .
-1- SIMILARITY: Belongs to the Afp-dependent AMP-binding enzyme
CC family.
DR EMBL; BX572598; CAE37228.1; -
DR GO; GO:0016874; F:ligase activity; IEA.
DR InterPro; IPR000873; AMP-bind.
DR Pfam; PF00501; AMP-binding; 1.
DR PRINTS; PR00154; AMPBINDING.
DR PROSITE; PS00455; AMP BINDING; 1.

```

Query Match	Score	DB 2	Length
Best Local Similarity	50.5%	Pred. No. 1.le-97	
Matches 292; Conservative	86;	Mismatches 154;	Indels 6; Gaps 2;

```

QY 1 MRSLEALLPPEGRILBLELHEMAKRRPROTCVAAANAAANGEMRRISYAEFHNVAIAQSL 60
Db 46 VR$TTLITDYFVRITRDLRHHFAETRAPEVFAERENGEGMRRISYAEMLRAAQITABALI 105
QY 61 PYGSAERPLIVSANGDLEHLOLAFGMAYGAIPIYCPVSPAYSLSDOLAKRHIVGLQOP 120
Db 106 ARGSAERPVWILSGNISIDHAMWFGALYAGVAMCVPSPSLVSKOYGUKLRHIVGLTP 165
QY 121 GLVFPADAPFORAIEITLLPDDVPAITRGGELAGRTVSPDSLLBQEGIEADNAFPAIG 180
Db 166 GLIFADDTITAFAPAILATVPEDVELATRGVKKRTYSLAELATPEHEPLAKHRAIG 225
QY 181 PDTIAKELPFGSGTKLEKAVPTTORMCANOMLOFPVPGEGEPVILWMLPNHNFEG 240
Db 226 HDITIAKELTSGSGNKPRAVINTORMCANOMVIREMAFLKDEPVIWMLPNHNRFGG 285
QY 241 SHNIGIYUNGCTYTLDDGKPTAQGFATLRLNLSIPTAYLTPPKMEELVGLAEBDST 300
Db 286 NHNIGILTFNGSGMYIDDGKPTAGIASTIRNLEINPTIYFNVPKYEISLLPYLRBDQ 345
QY 301 LRERFAPRMKLEFFAAAGLSQGIWDRDLRYABOHCGERIRMMAGLMTETAPSCFTTGP 360
Db 346 LRKLFESRLHMFPSGASLAHVWNGDEVAVAETGARVPMLTGLGATEPAFFPMSVTPQ 405
QY 361 LSMAGIYGLAPRGEVULPVDTGLEBRFHGPHMSGYMPAREONAAPEBEGYCGSDA 420
Db 406 T$RSGHGLVPGVGEAKLVPNNGLEVRAPGPNITPGIWRAPBELTIDAFPEBEGYKXDA 465
QY 421 IKLADPPADPOKGLMFDRIAEDFGLSGGFVSVPTLRTRAVLEGGSYVLDVVVAAPPREC 480
Db 466 LKRPDANDL$RGPFDPGRISBDFGLAAGTWSVGPLAKFI$AASLVROVVIAGLDRDY 525
QY 481 LGLLVFRLDL--CRAISGLGKEASDAEVLASEPVAWPAWMLKRLNREATYNASIRW 537
Db 526 VTALA---ILDPDGCULINATPLPLEDLAAGMAADHLIREAFRERPATLTLQATGSSNRVTR 582
QY 538 VGLLDPSPSIDKGEVTDKGSINQRAVUQMSAKDAILY 575
Db 583 AVLLGEPLSIDKGEITDKGSYNQCAVLEYEASLADLY 620

```

```

RESULT 12
CAE27228
ID CAE27228 PRELIMINARY; PRT; 632 AA.
AC CAE27228;
DT 02-MAR-2004 (TrEMBLrel. 27, Created)
DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT 02-MAR-2004 (TrEMBLrel. 27, Last annotation update)
DE Putative feruloyl-CoA synthetase (EC 6.2.1.-).
GN PCS2 OR RPA187.
OS Rhodopseudomonas palustris.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Rhodopseudomonas.
OX NCBI_TaxId:1076;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CGA009 / ATCC BAA-98;
RA PubMed=14704707;
RA Laimimer F.W., Chain P., Hauser L., Lamerdin J., Malfatti S., Do L.,
RA Land M.L., Pelleletter D.A., Beatty J.T., Lang A.S., Tabita F.R.,
RA Gibson J.L., Hanson T.E., Bobst C., Torres y Torres J.L., Perez C.,
RA Harrison F.H., Gibson J., Harwood C.S.;
RT "Complete genome sequence of the metabolically versatile
RT photosynthetic bacterium Rhodopseudomonas palustris.";
RL Nat. Biotechnol. 22:55-61(2004).
OR EMBL: BX572598; CAE27228.1; -

```

RC STRAIN=ADP1;
 RX MEDLINE=94237485; PubMed=8181753;
 RA Hartnett G.B., Ornston L.N.;
 RT "Acquisition of apparent DNA shippage structures during extensive
 RT evolutionary divergence of *pcaD* and *pcaD* genes encoding identical
 RT catalytic activities in *Acinetobacter calcoaceticus*.";
 RL Gene 142:23-29(1994).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ADP1;
 RX MEDLINE=94311565; PubMed=8063101;
 RA Kowalchuk G.A., Hartnett G.B., Benson A., Houghton J.E., Ngai K.L.,
 RA Ornston L.N.;
 RT "Contracting patterns of evolutionary divergence within the
 RT *Acinetobacter calcoaceticus* *pca* operon.";
 RL Gene 146:23-30(1994).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ADP1;
 RX MEDLINE=95095936; PubMed=8002591;
 RA Elismore D.A., Ornston L.N.;
 RT "The *pca* operon superoperonic cluster of *Acinetobacter calcoaceticus*
 RT contains *guaA*, the structural gene for *guanine*-shikimate
 RT dehydrogenase.";
 RL J. Bacteriol. 176:7659-7666(1994).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ADP1;
 RX MEDLINE=93194074; PubMed=8449410;
 RA Dimarco A.A., Averhoff B.A., Kim E.E., Ornston L.N.;
 RT "Evolutionary divergence of *pobA*, the structural gene encoding p-
 RT hydroxybenzoate hydroxylase in an *Acinetobacter calcoaceticus* strain
 RT well-suited for genetic analysis.";
 RL Gene 125:25-33(1993).
 RN [7]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ADP1;
 RX MEDLINE=93323239; PubMed=8331077;
 RA Dimarco A.A., Averhoff B., Ornston L.N.;
 RT "Identification of the transcriptional activator *pobR* and
 RT characterization of its role in the expression of *pobA*, the structural
 RT gene for p-hydroxybenzoate hydroxylase in *Acinetobacter*
 RT *calcoaceticus*.";
 RL J. Bacteriol. 175:4499-4506(1993).
 RN [8]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ADP1;
 RX MEDLINE=22401456; PubMed=12514037;
 RA Smith M.A., Weaver V.B., Young D.M., Ornston L.N.;
 RT "Genes for chlorogenate and hydroxycinnamate catabolism (*hca*) are
 RT linked to functionally related genes in the *dca-pca-gua* *pob-hca*
 RT chromosomal cluster of *Acinetobacter* sp. strain ADP1.";
 RL Appl. Environ. Microbiol. 69:524-532(2003).
 RN [9]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ADP1;
 RA Parke D., Ornston L.N.;
 RT "Hydroxycinnamate (*hca*) catabolic genes from *Acinetobacter* sp. strain
 RT ADP1 are repressed by HcaR and induced by hydroxycinnamoyl-CoA
 RT catabolism.";
 RL Appl. Environ. Microbiol. 0:0-0(2003).
 DR HSSP; L05770; AAL54850.1; -;
 DR HSSP; P08659; ILCT;
 DR GO; GO:0016874; F1; ligase activity; IEA.
 DR GO; GO:0008152; P; metabolism; IEA.
 DR InterPro; IPR000873; AMP-bind.
 DR Pfam; PF00501; AMP-binding; 1.
 KW ligase.
 SO SEQUENCE 626 AA; 70102 MW; 3C641E83ABDDC0DA CRC64;

Query Match 52.6%; Score 1621.5; DB 2; Length 626;
 Best Local Similarity 53.0%; Pred. No. 1e-109;
 Matches 308; Conservative 96; Mismatches 174; Indels 3; Gaps 3;

QY 5 BALPPGRILRLERLHMAKTRPEQTCVAAARAANGEMRRISYAEFHNVAIAOSLLPYGL 64
 DB 37 EQLKAYFQKLTDLHIFAKKRPYHTAAKRNTODEIRLSYAEITLQRAHIAQALQORNL 96
 QY 65 SAERPLIVSGNDLEHLOAFGAMVAGIPYCPVSPAYSLSDQLAKRHHVIGLLOGLVF 124
 DB 97 SQERPLIVISQNDLEHLLTSMGAMLAGVPFSAISPAYSLVSQDFGLKHVFDVLTGMYV 156
 QY 125 AADAAFPORAIETITLDDVPALFTFRELAGRRVPSDSLLEDPGGIEADNAAPATGPDT 184
 DB 157 ANDGAFAPAAIOSCTDHDHIEIVNKGIIIGQVCTSPQSLDTP-VTVDEHYETLDEHQI 215
 QY 185 AKFLPFGSGTKLPKAPVPTORMLCANQOMLOTFPVGSEPPLYMLPNNHFFGSHNI 244
 DB 216 AKFLPFGSGTKLPKAPVPTTHMLCINQOMLOTFPVEEDTLPVLLMLSHHFFGSHNV 275
 QY 245 GIVLYNGGTYYLDDGKPTAOGFAETLRNLSEISPTAYLTVPKMEELVGLERDSTLRER 304
 DB 276 GIALYNGGTYYLDDGKRPVQKFDETIRNLKEISPTYYLVNVPKMEELTDLERDAELRDR 335
 QY 305 FPARMKLPFPAAAGLSQGIWDRVAEQQCGRIMMAGLMTETAPSTFTTGPLSMA 364
 DB 336 FFANVKILFFAGALSEAGNNRLDKIAQACGKIRIMSGLMTETAPSCAFTTGRVMA 395
 QY 365 GYIGLPAPGCEVTLVPVQDLEGRFHPHYMSGVNR-APQONQAQFDEEGYCSGPAIKL 423
 DB 396 GFIGYAPGCEITLVYVQDLEGRFHPHYMSGVNR-APQONQAQFDEEGYCSGPAIKL 455
 QY 424 ADPADPQKLMFDPGRIAEDPFLSSGVFVSGPLRTAVLEGGSYLVLDVVAADRECLGL 483
 DB 456 VDBNDPAQGLMYDGRIAEDPFLKNTGTFVNVGTLRNKALIQGNLLIDVCTGSNLAIGF 515
 QY 484 LVPRRLDGRALSGLG-KEASDAEVLASEPVRVAFMDLKLRLNREKTNASRIMVGLLD 542
 DB 516 LVPRKLDACADFAGLMKRHSALDILQHPKVQGFQFLIYQNKDXTGSSNRVSMLYLMT 575
 QY 543 TPESIDKGEVTDKGSINORAVLQWRGSAKYDALYRGDSQML 583
 DB 576 EAPQLDAGEVTDKGNINQSNILKRAALVBEYLNKQTDNPL 616

RESULT 10
 Q6FBJ6 PRELIMINARY; PRT; 624 AA.
 AC Q6FBJ6;
 DT 05-JUL-2004 (TREMBLrel. 27, Created)
 DT 05-JUL-2004 (TREMBLrel. 27, last sequence update)
 DT 05-JUL-2004 (TREMBLrel. 27, last annotation update)
 DE Coenzyme A ligase (EC 6.2.1.3).
 GN Name=hcaC; OrderedAccession=AC1AD1724;
 OS *Acinetobacter* sp. (strain ADP1).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Moraxellaceae; *Acinetobacter*.
 OX NCBI_TaxID=62977;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Barbe V., Vallonet D., Fonknechten N., Kreimeyer A., Ortiz S.,
 RA Labarre V., Crivellier S., Robert C., Duprat S., Wincker P.,
 RA Ornston L.N., Weisenbach J., Marliere P., Cohen G.N., Medigue C.;
 RT "Unique features revealed by the genome sequence of *Acinetobacter* sp.
 RT ADP1, a versatile and naturally transformation competent bacterium.";
 RL Nucleic Acids Res. 0:0-0(2004).
 DR EMBL; CR543861; CAG68566.1; -;
 DR InterPro; IPR000873; AMP-bind.
 DR Pfam; PF00501; AMP-binding; 1.
 KW Complete proteome; Ligase.
 SO SEQUENCE 624 AA; 69856 MW; 724419C9BA506580 CRC64;

Query Match 52.5%; Score 1618.5; DB 2; Length 624;
 Best Local Similarity 52.8%; Pred. No. 1.7e-109;
 Matches 307; Conservative 97; Mismatches 174; Indels 3; Gaps 3;

QY 184 IAKFLFTSGSTKLPAKAVPTTORMLCANQOMLQTFPVFGSEPPVLVDMLPNNHTFGSSHN 243
 Db 217 IAKFLFTSGSTKLPAKAVTTQGMCANQOMLQTFPEFAEVPVLVDMLPNNHTFGSSHN 276
 QY 244 IGVLYNGGTYIYDDGKPTAQGFATLRNLSISPTLYLVTPKGMELVGALEPDSLTRE 303
 Db 277 VGVLYNGGTYIYDDGKPTAQGFATLRNLSISPTLYLVTPKGMELVGALEPDSLTRE 336
 QY 304 RFPARMKLPFFPAAGSQIWMRLDRVAEOHCGERIMMAGLWTEPTAPSCFTTGPLSM 363
 Db 337 RFPARMKLPFFPAAGSQIWMRLDRVAEOHCGERIMMAGLWTEPTAPSCFTTGPLSM 396
 QY 364 AGYIGLPAPCEVYLVDGKLEGRFPGHVMGSGYMPAPREOMAQAFDEGGYCSGDAIKL 423
 Db 397 AGYIGLPAPCEVYLVDGKLEGRFPGHVMGSGYMPAPREOMAQAFDEGGYCSGDAIKL 456
 QY 424 ADPAPOKGLMPDRIEDKLSGCVVSGPLTRAVLVBGGSYVLVVAAPRECLGL 483
 Db 457 CDEGNPQGLMPDRIEDKLSGCVVSGPLTRAVLVBGGSYVLVVAAPRECLGL 515
 QY 484 LVPEPRLDRCALSGLEKASDAEVLASEPYRAMFADMLKRLNREATGNASIMVVGILDT 543
 Db 516 LVPEPRLDRCALSGLEKASDAEVLASEPYRAMFADMLKRLNREATGNASIMVVGILDT 575
 QY 544 PPSIDKGEVTDKGSINQRAVLQWRSKAVDALYRG 577
 Db 576 APIDRGEITDKGSINQRAVLQWRSKAVDALYRG 609

RESULT 8

Q8XT88 PRELIMINARY; PRT; 624 AA.

ID Q8XT88
 AC Q8XT88
 DT 01-MAR-2002 (TREMblrel. 20, Created)
 DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)
 DT 01-MAR-2004 (TREMblrel. 26, Last annotation update)
 DE PROBABLE PERILOYL-COA SYNTHETASE PROTEIN.
 GN Name=fcf; Synonyms=RS05196; OrderedLocustNames=RS0227;
 OS Ralstonia solanacearum (Pseudomonas solanacearum).
 OC Plasmid megaplasmid.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Burkholderiaceae; Ralstonia.
 OX NCBI_TaxID=305;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GM1000;
 RX MEDLINE=21681879; PubMed=11823852;
 RA Salanoubat M., Genin S., Artiguenave F., Guzy J., Manganot S.,
 RA Atilat M., Billault A., Broctier P., Camus J.C., Cattoilico L.,
 RA Chandler M., Choisme N., Claudel-Renard C., Cunac S., Demange N.,
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiek T.,
 RA Siguler P., Thebaud P., Whalen M., Wincker P., Levy M.,
 RA Weissenbach J., Boucher C.A.
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum";
 RL Nature 415:497-502(2002).
 CC - - SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
 CC family.
 DR EMBL: AL646077; CAD17378.1; - -
 DR HSSP: P08659; ILCT.
 DR GO: GO:0003624; F: catalytic activity; IEA.
 DR GO: GO:0008152; P: metabolism; IEA.
 DR InterPro: IPR000873; AMP-bind.
 DR Pfam: PF00501; AMP-binding; 1.
 DR PRINTS: PR00154; AMPBINDING.
 KW Complete proteome; Plasmid.
 SQ SEQUENCE 624 AA; 67134 MW; 75EE2038ACFP25A1 CRC64;

Query Match 63.0%; Score 1940.5; DB 2; Length 624;
 Best Local Similarity 63.7%; Pred. NO. 5.1e-133;
 Matches 373; Conservative 69; Mismatches 143; Indels 1; Gaps 1;

QY 1 MRSLLEALPPGRILERLEHMAKTRPQTCVAAARAANGEWRRISYAEFHNVRALIAQSL 60

Db 35 LRSAPRIGAYBERMTDCLVNGAQOHPDRVLAARGADGQMERITYAQMHRARAVGALL 94
 QY 61 PYGSAERPLIYSGNDLEHLOLAFGAMVAGIYPCPVSPVYSLSDPLAKLRHVGLOP 120
 Db 95 ARGISPERPLIISGNDLQHLQALGAMVAGIYPCPVSPVYSLSDPLAKLRHVGLOP 154
 QY 121 GLVPAADAPFORAETILDPDPAIFTRGELAGRTVSPSLLEOPGIEADNAPATG 180
 Db 155 GLVYATGALFAGALQAVVPEPETEVALIDHGEVAGRAVTRIASLI-ATGPADVDAANARVG 213
 QY 181 PDIITAKLFTSGSTKLPAKAVPTTORMLCANQOMLQTFPVFGSEPPVLVDMLPNNHTFGG 240
 Db 214 PDIITAKLFTSGSTKSPKATTTIRMLCSNQOMLQTFPEFGAPVLLDMLPNNHTFGG 273
 QY 241 SHNIGIYLYNGGTYIYDDGKPTAQGFATLRNLSISPTLYLVTPKGMELVGALEPDSLTRE 300
 Db 274 SHNIGIYLYNGGTYIYDDGKPTAQGFATLRNLSISPTLYLVTPKGMELVGALEPDSLTRE 333
 QY 301 LRERFARMKLPFFPAAGSQIWMRLDRVAEOHCGERIMMAGLWTEPTAPSCFTTGP 360
 Db 334 LRERFARMKLPFFPAAGSQIWMRLDRVAEOHCGERIMMAGLWTEPTAPSCFTTGP 393
 QY 361 LSNAGTYGLPAPCEVYLVDGKLEGRFPGHVMGSGYMPAPREOMAQAFDEGGYCSGDA 420
 Db 394 LSNAGTYGLPAPCEVYLVDGKLEGRFPGHVMGSGYMPAPREOMAQAFDEGGYCSGDA 453
 QY 421 IXLADPADPOKGLMPDRIEDKLSGCVVSGPLTRAVLVBGGSYVLVVAAPRECLGL 480
 Db 454 IXLADPADPOKGLMPDRIEDKLSGCVVSGPLTRAVLVBGGSYVLVVAAPRECLGL 513
 QY 481 LGLVPEPRLDRCALSGLEKASDAEVLASEPYRAMFADMLKRLNREATGNASIMVVGILDT 540
 Db 514 LGLVPEPRLDRCALSGLEKASDAEVLASEPYRAMFADMLKRLNREATGNASIMVVGILDT 573
 QY 541 LDTPEPIDKGEVTDKGSINQRAVLQWRSKAVDALYRGEDSMRLDE 586
 Db 574 LDTPEPIDKGEVTDKGSINQRAVLQWRSKAVDALYRGEDSMRLDE 619

RESULT 9

Q8RM01 PRELIMINARY; PRT; 626 AA.

ID Q8RM01
 AC Q8RM01
 DT 01-JUN-2002 (TREMblrel. 21, Created)
 DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
 DT 01-JUN-2003 (TREMblrel. 24, Last annotation update)
 DE Coenzyme A ligase.
 GN Name=hcaC;
 OS Acinetobacter sp. (strain ADP1).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Moraxellaceae; Acinetobacter.
 OX NCBI_TaxID=62977;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ADP1;
 RX MEDLINE=98175676; PubMed=9515921;
 RA Gerischer U., Segura A., Ornston L.N.;
 RT "Peau, a transcriptional activator of genes for protocatechuate
 RL utilization in Acinetobacter";
 RN J. Bacteriol. 180:1512-1524(1998).
 RC SEQUENCE FROM N.A.
 RC STRAIN=ADP1;
 RX MEDLINE=90130333; PubMed=2298704;
 RA Hartnett C., Neidle E.L., Ngai K.L., Ornston L.N.;
 RT "DNA sequences of genes encoding Acinetobacter calcoaceticus
 RT protocatechuate 3,4-dioxygenase: evidence indicating shuffling of
 RT genes and of DNA sequences within genes during their evolutionary
 RL divergence";
 RN J. Bacteriol. 172:956-966(1990).
 RP [3]
 RP SEQUENCE FROM N.A.

```
QY 61 PYGLSAERPLILVSGNDLEHLQALFAGMAYAGIPYCVSPAYSLLSODLAKRHIVQTLQPGYV 120
DB 61 GLGLSABERPLALISGNDLEHLQALFAGMAYAGIYCVSPAYSLLSODPALRHVCEVLTP 120
QY 121 GLVFADAAAPFORAIEITLIPDVPAIFTRGELAGRRTVSFDSLQPGGI- EADNAFAAT 179
DB 121 GVFVPSDSQPFQAFAPAVLDSDSGVSVNGQVAGRHHISDSL- QPGDLAADAFAAT 179
QY 180 GPDITAKFLFTSGSTYLPKAVVTTQRMLCANQOMLQTFEPVGESEPPVLVDMLPMNHTTG 239
DB 180 GPDITAKFLFTSGSTYLPKAVVTTQRMLCANQOMLQTFEPVGESEPPVLVDMLPMNHTTG 239
QY 240 GSHNIGIVLYNGSTYLLDQKPTAOGFAETRLNLSISPTAYLTVPKGMEVLGALERS 239
DB 240 GSHNIGIVLYNGSTYLLDQKPTAOGFAETRLNLSISPTAYLTVPKGMEVLGALERS 239
QY 300 TLREFFAAMKLFYFAAGLSQGIWDLRDLVAHQHGERIRMMAGLGMTETASCTFTTG 359
DB 300 ALREVEFAIILKLFYFAAGLSQSVMDRLDLVAHQHGERIRMMAGLGMTETASCTFTTG 359
QY 360 PLSMAGYIGLPAGCEYKLVVDGKLEGRFPHVMSGYRABEQNAQAFDEGGYCSGD 419
DB 360 PLSMAGYIGLPAGCEYKLVVDGKLEGRFPHVMSGYRABEQNAQAFDEGGYCSGD 419
QY 420 AITKLADPADPOKGLMFDGRIADPFKLSGCVFVSGPLRTRAVLGGSYLDDVVVAAPDRE 479
DB 420 ALKLADARQEBGLMFDGRIADPFKLSGCVFVSGPLRTRAVLGGSYLDDVVVAAPDRE 479
QY 480 CIGLLVFPRLDGRALSGLKEASDAEVLASEPVRAMPADWLKRLREATGNASRIIMWG 539
DB 480 CIGLLVFPRLDGRALSGLKEASDAEVLASEPVRAMPADWLKRLREATGNASRIIMWG 539
QY 540 LLDTPPSINKGEVTDKGSINQRAVLOWRSAKVDALYRGDSML 583
DB 540 LLDTPPSINKGEVTDKGSINQRAVLOWRSAKVDALYRGDSML 583
```

RESULT 6

```
Q70V06 PRELIMINARY; PRT; 618 AA.
ID Q70V06;
AC Q70V06;
DT 05-JUL-2004 (T-EMBLrel. 27, Created)
DT 05-JUL-2004 (T-EMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (T-EMBLrel. 27, Last annotation update)
DE Feruloyl-CoA synthetase.
GN Name=fcS;
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=303;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=T2;
RA Ficca A.G., Di Gioia D., Barghini P., Fava F., Ruzzi M.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
CC -1- Similarity: Belongs to the ATP-dependent AMP-binding enzyme
family.
DR EMBL: AU536324; CAD60263.1; -.
DR InterPro: IPR000873; AMP-bind.
DR Pfam: PF00501; AMP-binding.1.
DR PRINTS: PR00154; AMPBINDING.
SQ SEQUENCE 618 AA; 66992 MW; 57EEEB41166EBD33 CRC64;
```

Query Match 68.2%; Score 2103; DB 2; Length 618;
Best Local Similarity 71.4%; Pred. No. 6.8e-145;
Matches 410; Conservative 48; Mismatches 114; Indels 2; Gaps 2;

```
QY 4 LEALLFPGRILRLHMAKTRPEQTCVABANGERRISYEMFHNVAIAQSLLPYG 63
DB 38 VETLDYVPAALMRLHMAQORSEQTFVARRADGQWQHITVAAMLQVRRIATWLLGQG 97
QY 64 LSAERPLILVSGNDLEHLQALFAGMAYAGIPYCVSPAYSLLSODLAKRHIVQTLQPGYV 123
```

```
DB 98 LSAERPLILVSGNDLEHLQALFAGMAYAGIPYCVSPAYSLLSODFGKLRIHVQTLQPGYV 157
QY 124 FAADAAAPFORAIEITLIPDVPAIFTRGELAGRRTVSFDSLQPGGI- EADNAFAATGPT 183
DB 158 FAADAA- YGRAIDAVEEPSINIVLGGGELAGNASTPFAAVLTTVAQPCADQAFALRPDS 216
QY 184 IAKFLFTSGSTYLPKAVVTTQRMLCANQOMLQTFEPVGESEPPVLVDMLPMNHTTFCGSHN 243
DB 217 LAKFLFTSGSTYLPKAVVTTQRMLCANQOMLQTFEPVGESEPPVLVDMLPMNHTTFCGSHN 276
QY 244 IGIVLYNGSTYLLDQKPTAOGFAETRLNLSISPTAYLTVPKGMEVLGALERS 303
DB 277 VGIVLYNGSTYLLDQKPTAOGFAETRLNLSISPTAYLTVPKGMEVLGALERS 336
QY 304 RFPARMKLFYFAAGLSQGIWDLRDLVAHQHGERIRMMAGLGMTETASCTFTTG 363
DB 337 RFPARMKLFYFAAGLSQGIWDLRDLVAHQHGERIRMMAGLGMTETASCTFTTG 396
QY 364 AGYIGLPAGCEYKLVVDGKLEGRFPHVMSGYRABEQNAQAFDEGGYCSGDAILK 423
DB 397 AGYIGLPAGCEYKLVVDGKLEGRFPHVMSGYRABEQNAQAFDEGGYCSGDAILK 456
QY 424 ADPADPOKGLMFDGRIADPFKLSGCVFVSGPLRTRAVLGGSYLDDVVVAAPDRECLG 483
DB 457 CDRGNPQLGIMFDGRIADPFKLSGCVFVSGPLRTRAVLGGSYLDDVVVAAPDRECLG 515
QY 484 LVFPRLDGRALSGLKEASDAEVLASEPVRAMPADWLKRLREATGNASRIIMWG 543
DB 516 LVFPRLDGRALSGLKEASDAEVLASEPVRAMPADWLKRLREATGNASRIIMWG 575
QY 544 PSIDKGEVTDKGSINQRAVLOWRSAKVDALYRG 577
DB 576 APIDNGEITDKGSINQRAVLOWRSAKVDALYRG 609
```

RESULT 7

```
Q70V06 PRELIMINARY; PRT; 618 AA.
ID CAD60263;
AC CAD60263;
DT 02-MAR-2004 (T-EMBLrel. 27, Created)
DT 02-MAR-2004 (T-EMBLrel. 27, Last sequence update)
DT 02-MAR-2004 (T-EMBLrel. 27, Last annotation update)
DE Feruloyl-CoA synthetase.
GN Name=fcS;
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=303;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=T2;
RA Ficca A.G., Di Gioia D., Barghini P., Fava F., Ruzzi M.;
RL "Identification of Pseudomonas putida strain T2 genes involved in
ferulate catabolism.";
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL: AU536324; CAD60263.1; -.
SQ SEQUENCE 618 AA; 66992 MW; 57EEEB41166EBD33 CRC64;
```

Query Match 68.2%; Score 2103; DB 2; Length 618;
Best Local Similarity 71.4%; Pred. No. 6.8e-145;
Matches 410; Conservative 48; Mismatches 114; Indels 2; Gaps 2;

```
QY 4 LEALLFPGRILRLHMAKTRPEQTCVABANGERRISYEMFHNVAIAQSLLPYG 63
DB 38 VETLDYVPAALMRLHMAQORSEQTFVARRADGQWQHITVAAMLQVRRIATWLLGQG 97
QY 64 LSAERPLILVSGNDLEHLQALFAGMAYAGIPYCVSPAYSLLSODLAKRHIVQTLQPGYV 123
DB 98 LSAERPLILVSGNDLEHLQALFAGMAYAGIPYCVSPAYSLLSODPKLRIHVQTLQPGYV 157
QY 124 FAADAAAPFORAIEITLIPDVPAIFTRGELAGRRTVSFDSLQPGGI- EADNAFAATGPT 183
DB 158 FAADAA- YGRAIDAVEEPSINIVLGGGELAGNASTPFAAVLTTVAQPCADQAFALRPDS 216
```

```

QY 181 PPTIAKFLFTSGSTKLPKAVPTTORMLCANQOMLQTFPFVGESEPPVLVDMLPMNHTFGG 240
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 PSIAKFLFTSGSTKLPKAVITTTQRMLCANQOMLQTFPFVGESEPPVLVDMLPMNHTFGG 240
QY 241 SHNIGIVLYNGGTYIYDDGKPTAQGFATLRNLSEISPTAYLTVPKGMEELVGALEBDST 300
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 241 SHNVGIVLYNGGTFYVDEKGPPTAQGFATLRNLKEISPTAYLTVPKGMEELVVALLEQDAE 300
QY 301 LNERFPAARKLFFPAAAGSOGIWDRLDVAHQCEERIRMGAGLMTETAPSCFTTGP 360
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 301 LNERCFKRNLSLFFPAAAGSOSQSWDRLDVAHQCEERIRMGAGLMTETAPSCFTTGP 360
QY 361 LSMAGYIGPAPCEVKLVVDGKLEGRFPHGPHVMSGYRARPEQNAQAFDEEYSCGDA 420
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 361 LSMAGYIGPAPCEVKLVVDGKLEGRFPHGPHVMSGYRARPEQNAQOTAEVDAQYFCSGDA 420
QY 421 IKLADPADPQKGMFDGRIAEDEFKLSGGVFSVGPILRTAVIEGSSYVLDDVVAADPDR 480
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 421 IKLADPADPQKGMFDGRIAEDEFKLSGGVFSVGPILRTAVIEGSSYVLDDVVAADPDR 480
QY 481 LGLVFPRLDGRALSGLGKEASDAVLASEPVPAPADMLKRLNEATGNASRIMWGL 540
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 481 LGLVFPRLDGRALSGLGKEASDAVLASEPVPAPADMLKRLNEATGNASRIMWGL 540
QY 541 LDTPPSIDKGEVTDKGSINQRAVLQWRSKAVDALYRGEDQSMRLDE 586
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 541 QVEBPASIDGGEITDKGSINQRAVLQWRAQVENVYRGREPSILRAE 586

RESULT 4
CAD60268 PRELIMINARY; PRT; 589 AA.
ID CAD60268
AC CAD60268
DT 02-MAR-2004 (TEMBLrel. 27, Created)
DT 02-MAR-2004 (TEMBLrel. 27, Last sequence update)
DE Feruloyl-CoA synthetase.
GN FCS.
OS Pseudomonas fluorescens.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=294;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BF13.
RA Barchini P., Civolani C., Ficca A.G., Schieser A., Ruzzi M.;
RT "Cloning and Characterization of Ferulate Catabolism Genes from
RT Pseudomonas fluorescens BF13."
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ536325; CAD60268.1; -.
SQ SEQUENCE 589 AA; 64311 MW; ECDC7DA0EDD5552B CRC64;

Query Match 79.5%; Score 2449; DB 2; Length 589;
Best Local Similarity 77.8%; Pred. No. 3.4e-170;
Matches 456; Conservative 54; Mismatches 76; Indels 0; Gaps 0;

QY 1 MSLTEALLFPGRILRLTHMAKTRPEQTCVBARANGEMRRISYVEMFNVAIAIOSLL 60
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MSLTKLALPDLRLVHMAVREQTFIARREGSGMRLVSYQMDLSVAIAIOGLL 60
QY 61 PGLSARPLLVSGNDLEHLQALFAGMAGTIPYCVSPAYSLLSODLAKLRHIVGLDP 120
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 RYGLSMDKRLALLSSGNDIEHLQALGAMTAGTIPYCVSPAYSLLSODPAKLRHIVCOLDP 120
QY 121 GLVPAADAPFORALETILPDDVPAIFTRGELAGRTTVPFSDILLQPGGIEADNAPATG 180
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 GLVFSADASAYORADAVLPETPLISVGRQVGRQASFAISLAPGAEADAPAAATG 180
QY 181 PPTIAKFLFTSGSTKLPKAVPTTORMLCANQOMLQTFPFVGESEPPVLVDMLPMNHTFGG 240
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 PSIAKFLFTSGSTKLPKAVITTTQRMLCANQOMLQTFPFVGESEPPVLVDMLPMNHTFGG 240

```

```

QY 241 SHNIGIVLYNGGTYIYDDGKPTAQGFATLRNLSEISPTAYLTVPKGMEELVGALEBDST 300
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 241 SHNVGIVLYNGGTFYVDEKGPPTAQGFATLRNLKEISPTAYLTVPKGMEELVVALLEQDAE 300
QY 301 LNERFPAARKLFFPAAAGSOGIWDRLDVAHQCEERIRMGAGLMTETAPSCFTTGP 360
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 301 LNERCFKRNLSLFFPAAAGSOSQSWDRLDVAHQCEERIRMGAGLMTETAPSCFTTGP 360
QY 361 LSMAGYIGPAPCEVKLVVDGKLEGRFPHGPHVMSGYRARPEQNAQAFDEEYSCGDA 420
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 361 LSMAGYIGPAPCEVKLVVDGKLEGRFPHGPHVMSGYRARPEQNAQOTAEVDAQYFCSGDA 420
QY 421 IKLADPADPQKGMFDGRIAEDEFKLSGGVFSVGPILRTAVIEGSSYVLDDVVAADPDR 480
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 421 IKLADPADPQKGMFDGRIAEDEFKLSGGVFSVGPILRTAVIEGSSYVLDDVVAADPDR 480
QY 481 LGLVFPRLDGRALSGLGKEASDAVLASEPVPAPADMLKRLNEATGNASRIMWGL 540
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 481 LGLVFPRLDGRALSGLGKEASDAVLASEPVPAPADMLKRLNEATGNASRIMWGL 540
QY 541 LDTPPSIDKGEVTDKGSINQRAVLQWRSKAVDALYRGEDQSMRLDE 586
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 541 QVEBPASIDGGEITDKGSINQRAVLQWRAQVENVYRGREPSILRAE 586

RESULT 5
Q88HKO PRELIMINARY; PRT; 589 AA.
ID Q88HKO
AC Q88HKO
DT 01-JUN-2003 (TEMBLrel. 24, Created)
DT 01-JUN-2003 (TEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TEMBLrel. 26, Last annotation update)
DE Feruloyl-CoA synthetase.
GN Name=fcg; OrderedlocusNames=PP3356;
OS Pseudomonas putida (strain KT2440).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=160488;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22423060; PubMed=12534463;
RA Nelson K.E., Weinel C., Paulsen I.T., Dodson R.J., Hilbert H.,
RA Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Popp M., Holmes M.,
RA Brinkac L.M., Beanan M.J., DeBoy R.T., Daugherty S.C., Kolonay J.F.,
RA Madupu R., Nelson W.C., White O., Peterson J.D., Khouri H.M.,
RA Hance I., Chris Lee P., Holtzapple E.K., Scanlan D., Tran K.,
RA Moazzez A., Uterback T.R., Rizzo M., Lee K., Kosack D., Moestl D.,
RA Wedler H., Lauber J., Stjepandic D., Hohnselt U., Stracz M., Heim S.,
RA Kiewitz C., Eisen J.A., Timmis K.N., Duesterhoeft A., Tuenmler B.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative analysis of the
RT metabolically versatile Pseudomonas putida KT2440."
RL Environ. Microbiol. 4:799-808(2002).
CC -1- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
CC family.
DR EMBL; AB016786; AAN68960.1; -.
DR TIGR; PP3356; -.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR000873; AMP-bind.
DR Pfam; PF00501; AMP-binding; 1.
DR PRINTS; PR00154; AMPBINDING.
KW Complete proteome.
SQ SEQUENCE 589 AA; 64283 MW; 4B02BCB47253237 CRC64;

Query Match 76.2%; Score 2349; DB 2; Length 589;
Best Local Similarity 75.9%; Pred. No. 6.9e-163;
Matches 443; Conservative 56; Mismatches 83; Indels 2; Gaps 2;

QY 1 MSLTEALLFPGRILRLTHMAKTRPEQTCVBARANGEMRRISYVEMFNVAIAIOSLL 60
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MQVEPLALPDLRLVHMAQVRPDTTFIARQDAGMRISISYVQMLADVATIANLL 60

```

```

QY 361 LSMAGYIGLPAPGCEVKLVVDGKLEGRFHGHVMSGYRARPONAOAEDBEGYCSGA 420
DB 361 LSMAGYIGLPAPGCEVKLVVDGKLEGRFHGHVMSGYRARPONAOAEDBEGYCSGA 420
QY 421 IKLADPADPQKGLMFGDRIAEDEPKLSSGVFSVGPLRTAVLEGSGYLDVVAAPDREC 480
DB 421 IKLADPADPQKGLMFGDRIAEDEPKLSSGVFSVGPLRTAVLEGSGYLDVVAAPDREC 480
QY 481 LGLVFPRLDRCALSGLGKEASDAEVLASEPVRAFADMLKRLNREATGNASRIWVGL 540
DB 481 LGLVFPRLDRCALSGLGKEASDAEVLASEPVRAFADMLKRLNREATGNASRIWVGL 540
QY 541 LDTPEIDGSEVTDKGSINQRAVLQWRSAKVDAALYGEQDSMLRDEATL 589
DB 541 LDTPEIDGSEVTDKGSINQRAVLQWRSAKVDAALYGEQDSMLRDEATL 589

```

RESULT 2

```

Q881P2 PRELIMINARY; PRT; 589 AA.
ID 0881P2;
AC 0881P2;
DT 01-JUN-2003 (TEMBLrel. 24, Created)
DT 01-JUN-2003 (TEMBLrel. 24, Last sequence update)
DE 01-MAR-2004 (TEMBLrel. 26, Last annotation update)
DE Feruloyl-CoA synthetase.
GN Name=fcsl; OrderedLocuNames=PSPT02941;
OS Pseudomonas syringae (pv. tomato).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxId=323;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RX MEDLINE=22834015; PubMed=12928499; DOI=10.1073/pnas.1731982100;
RA Buehl C.R., Joardar V., Lindeberg M., Selengut J., Paulsen I.T.,
RA Gwin M.L., Dodson R.J., Deboy R.T., Durkin A.S., Kolonay J.F.,
RA Madupu R.C., Daugherty S.C., Brinkac L.M., Beaman M.J., Haft D.H.,
RA Nelson W.C., Davidson T.M., Tran B., Russell D., Berry K.U.,
RA Khouri H.M., Fedorova N.B., Tran B., Russell D., Berry K.U.,
RA Utterback T.R., Van Aken S.E., Feldblyum T.V., D'Ascenzo M.,
RA Deng W.-L., Ramos A.R., Alfano J.R., Cartinhour S., Chatterjee A.K.,
RA Delaney T.P., Lazarowitz S.G., Martin G.B., Schneider D.J., Tang X.,
RA Bender C.L., White O., Fraser C.M., Collier A.;
RT "The complete genome sequence of the Arabidopsis and tomato pathogen
RT Pseudomonas syringae pv. tomato DC3000."
RT Proc. Natl. Acad. Sci. U.S.A. 100:10181-10186(2003).
CC -1- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
CC family.
DR EMBL: AE016866; AAO56434.1; -.
DR HSSP: P08659; 1LCT.
DR TIGR: PSP02941; -.
DR GO: GO:0003824; F: catalytic activity; IEA.
DR GO: GO:0008152; P: metabolism; IEA.
DR InterPro: IPR000873; AMP-bind.
DR Pfam: PF00501; AMP-binding; 1.
DR PRINTS: PR00154; AMPBINDING.
KW Complete proteome.
SQ SEQUENCE 589 AA; 64272 MW; FA135BCB9722AF12 CRC64;

```

```

Query Match 80.3%; Score 2474; DB 2; Length 589;
Best Local Similarity 78.9%; Pred. No. 5, 1e-172;
Matches 460; Conservative 50; Mismatches 73; Indels 0; Gaps 0;
QY 1 MRSLEALLPFGGILRLLEHMAKTRPEQTCVAAARAANGERRISYAEFHNVAIAOSLL 60
DB 1 MRSLEALLPFGGILRLLEHMAKTRPEQTCVAAARAANGERRISYAEFHNVAIAOSLL 60
QY 61 PYGLSARPLLIVSGNDLHLQALFAGNAGIPYCVSPAYSLSDOLAKRIIVGLTOP 120
DB 61 PYGLSARPLLIVSGNDLHLQALFAGNAGIPYCVSPAYSLSDOLAKRIIVGLTOP 120
QY 61 PYGLSARPLLIVSGNDLHLQALFAGNAGIPYCVSPAYSLSDOLAKRIIVGLTOP 120
DB 61 PYGLSARPLLIVSGNDLHLQALFAGNAGIPYCVSPAYSLSDOLAKRIIVGLTOP 120
QY 121 GLVFADAAAPFORAIFETILPDVPAIFTRGELAGRTVSFDSLLEQPGIEADNAFAATG 180
DB 121 GLVFADAAAPFORAIFETILPDVPAIFTRGELAGRTVSFDSLLEQPGIEADNAFAATG 180

```

```

DB 121 GLVFADAAAPFORAIFETILPDVPAIFTRGELAGRTVSFDSLLEQPGIEADNAFAATG 180
QY 181 PDIAKFLFTSGSTKLPKAVPTTORMLCANQOMLOTFFVGESEPPVLDMLPMNHTFGG 240
DB 181 PDIAKFLFTSGSTKLPKAVPTTORMLCANQOMLOTFFVGESEPPVLDMLPMNHTFGG 240
QY 241 SHNIGIVLVNGGIVYLDGDKPTAGFAETLRNLSISPAVLYLVPKGMELVGLRRDST 300
DB 241 SHNIGIVLVNGGIVYLDGDKPTAGFAETLRNLSISPAVLYLVPKGMELVGLRRDST 300
QY 301 LRRFPFARKMLPFPAAGSGOGIWDRLDVAEHCGERIRMAAGLQMTETAPSCFTTGP 360
DB 301 LRRFPFARKMLPFPAAGSGOGIWDRLDVAEHCGERIRMAAGLQMTETAPSCFTTGP 360
QY 361 LSMAGYIGLPAPGCEVKLVVDGKLEGRFHGHVMSGYRARPONAOAEDBEGYCSGA 420
DB 361 LSMAGYIGLPAPGCEVKLVVDGKLEGRFHGHVMSGYRARPONAOAEDBEGYCSGA 420
QY 421 IKLADPADPQKGLMFGDRIAEDEPKLSSGVFSVGPLRTAVLEGSGYLDVVAAPDREC 480
DB 421 IKLADPADPQKGLMFGDRIAEDEPKLSSGVFSVGPLRTAVLEGSGYLDVVAAPDREC 480
QY 481 LGLVFPRLDRCALSGLGKEASDAEVLASEPVRAFADMLKRLNREATGNASRIWVGL 540
DB 481 LGLVFPRLDRCALSGLGKEASDAEVLASEPVRAFADMLKRLNREATGNASRIWVGL 540
QY 541 LDTPEIDGSEVTDKGSINQRAVLQWRSAKVDAALYGEQDSML 583
DB 541 LDTPEIDGSEVTDKGSINQRAVLQWRSAKVDAALYGEQDSML 583

```

RESULT 3

```

Q70V01 PRELIMINARY; PRT; 589 AA.
ID Q70V01;
AC Q70V01;
DT 05-JUL-2004 (TEMBLrel. 27, Created)
DT 05-JUL-2004 (TEMBLrel. 27, Last sequence update)
DE 05-JUL-2004 (TEMBLrel. 27, Last annotation update)
DE Feruloyl-CoA synthetase.
GN Name=fcsl;
OS Pseudomonas fluorescens.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxId=294;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BF13;
RA Barghini P., Civolani C., Ficca A.G., Schiesser A., Ruzzi M.;
RT Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
CC family.
DR EMBL: AU536325; CAD60268.1; -.
DR InterPro: IPR000873; AMP-bind.
DR Pfam: PF00501; AMP-binding; 1.
DR PRINTS: PR00154; AMPBINDING.
SQ SEQUENCE 589 AA; 64311 MW; ECDCTD40BDD55F2B CRC64;

```

```

Query Match 79.5%; Score 2449; DB 2; Length 589;
Best Local Similarity 77.8%; Pred. No. 3, 4e-170;
Matches 456; Conservative 54; Mismatches 76; Indels 0; Gaps 0;
QY 1 MRSLEALLPFGGILRLLEHMAKTRPEQTCVAAARAANGERRISYAEFHNVAIAOSLL 60
DB 1 MRSLEALLPFGGILRLLEHMAKTRPEQTCVAAARAANGERRISYAEFHNVAIAOSLL 60
QY 61 PYGLSARPLLIVSGNDLHLQALFAGNAGIPYCVSPAYSLSDOLAKRIIVGLTOP 120
DB 61 PYGLSARPLLIVSGNDLHLQALFAGNAGIPYCVSPAYSLSDOLAKRIIVGLTOP 120
QY 61 PYGLSARPLLIVSGNDLHLQALFAGNAGIPYCVSPAYSLSDOLAKRIIVGLTOP 120
DB 61 PYGLSARPLLIVSGNDLHLQALFAGNAGIPYCVSPAYSLSDOLAKRIIVGLTOP 120
QY 121 GLVFADAAAPFORAIFETILPDVPAIFTRGELAGRTVSFDSLLEQPGIEADNAFAATG 180
DB 121 GLVFADAAAPFORAIFETILPDVPAIFTRGELAGRTVSFDSLLEQPGIEADNAFAATG 180

```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2005, 21:38:30 ; Search time 138 Seconds
(without alignments)
2455.766 Million cell updates/sec

Title: US-09-750-986D-30
Perfect score: 3082
Sequence: 1 MSLEALPPGRILRLERH.....KYDALYRGEDQSMRLDEATL 589

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues
Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_02:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3082	100.0	589	2	Q9RLD9 pseudomonas
2	2474	80.3	589	2	Q881F2 pseudomonas
3	2449	79.5	589	2	Q70V01 pseudomonas
4	2449	79.5	589	2	CAD60268 pseudomon
5	2349	76.2	589	2	Q88HK0 pseudomonas
6	2103	68.2	618	2	Q70V06 pseudomonas
7	2103	68.2	618	2	CAD60263 pseudomon
8	1940.5	63.0	624	2	Q8XT88 ralsstonia s
9	1621.5	52.6	626	2	Q8RM01 acinetobact
10	1618.5	52.5	624	2	Q6FBJ6 acinetobact
11	1457	47.3	632	2	Q6N8W6 pseudomonas
12	1457	47.3	632	2	CAE27228 rhodosphe
13	1401	45.5	627	2	Q89CF2 pseudomonas
14	1294	42.0	591	2	Q7CZ88 agrobacteri
15	1294	42.0	611	2	Q8UF11 agrobacteri
16	1210.5	39.3	644	2	Q89VJ9 bradyrhizob
17	1129	36.6	623	2	Q89CJ5 bradyrhizob
18	1128	36.6	624	2	Q6N8Y3 rhodopseu
19	1128	36.6	624	2	CAE27148 rhodopseu
20	1002.5	32.5	596	2	Q9AB43 caulobacter
21	972	31.5	621	2	Q9A368 pseudomonas
22	972	31.5	621	2	Q93X79 pseudomonas
23	972	31.5	621	2	Q93X79 pseudomonas
24	972	31.5	621	2	Q93X79 pseudomonas
25	972	31.5	621	2	Q93X79 pseudomonas
26	972	31.5	621	2	Q93X79 pseudomonas
27	972	31.5	621	2	Q93X79 pseudomonas
28	972	31.5	621	2	Q93X79 pseudomonas
29	972	31.5	621	2	Q93X79 pseudomonas
30	972	31.5	621	2	Q93X79 pseudomonas
31	972	31.5	621	2	Q93X79 pseudomonas

32	289	9.4	601	2	AAS63793 yersinia
33	289	9.4	619	2	Q89UX1 bradyrhizob
34	288	9.3	658	2	Q8DJC6 synechococc
35	286	9.3	607	1	LCFH_HAMIN
36	284	9.2	700	2	Q8LKS5 arabidopsis
37	283.5	9.2	601	2	Q8AIV8 bacteroides
38	283.5	9.2	601	2	Q9KPB9 vibrio chol
39	281	9.1	612	2	Q82UE1 nitrosomona
40	281	9.1	701	2	Q9CSU7 arabidopsis
41	280.5	9.1	602	2	Q87S81 vibrio para
42	280	9.1	604	2	Q6LV19 photobacter
43	280	9.1	604	2	CAG18856 photobact
44	279.5	9.1	604	2	Q73WT0 mycobacteri
45	279.5	9.1	473	2	AAS04897 mycobacte

ALIGNMENTS

RESULT 1	ID	Q9RLD9	PRELIMINARY;	PRT;	589 AA.
AC	Q9RLD9				
DT	01-MAY-2000	(TREMBLrel. 13, Created)			
DT	01-MAY-2000	(TREMBLrel. 13, last sequence update)			
DT	01-OCT-2003	(TREMBLrel. 25, last annotation update)			
DE	Penicillin-CoA synthetase.				
GN	Name=fcgs;				
OS	Pseudomonas sp.				
OC	Bacteria; Proteobacteria.				
OX	NCBI_TaxID=306;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=HR199;				
RX	MEDLINE=20011220; PubMed=10543794;				
RA	Overhage J., Frieleert H., Steinduechel A.,				
RT	"Biochemical and genetic analyses of the ferulic acid catabolism in				
RT	Pseudomonas sp. strain HR199."				
RL	Appl. Environ. Microbiol. 65:4837-4847(1999).				
DR	EMBL: AJ28746; CAB6026.1; -.				
DR	GO: GO:0003824; F: catalytic activity; IEA.				
DR	GO: GO:0008152; P: metabolic; IEA.				
DR	InterPro: IPR000873; AMP-bind.				
DR	Pfam: PF0501; AMP-binding; 1.				
SQ	SEQUENCE 589 AA; 64512 MW; 60C9A9B6AF78BCD CRC64;				

Query Match 100.0%; Score 3082; DB 2; Length 589;
Best Local Similarity 100.0%; Pred. No. 2e-216;
Matches 589; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MSLEALPPGRILRLERHMAKTRPEOTCVARAANGERRISYAEPMHNVAIAOSL	60
DB	1	MSLEALPPGRILRLERHMAKTRPEOTCVARAANGERRISYAEPMHNVAIAOSL	60
QY	61	PYGLSARPLLIVSGNDLRLQLAFGAMYAGIPYCVSPAYSLISODLAKRHIVGLLP	120
DB	61	PYGLSARPLLIVSGNDLRLQLAFGAMYAGIPYCVSPAYSLISODLAKRHIVGLLP	120
QY	121	GLVFADAAAPQPAITIIIPDVPAIFTRGELAGRTVSFDSLEOPGIEADNAPATG	180
DB	121	GLVFADAAAPQPAITIIIPDVPAIFTRGELAGRTVSFDSLEOPGIEADNAPATG	180
QY	181	PDTIAKFLFTSGSTKLPKAVPTTQRMLCANQQLQTFVPFGSEPPVLVDWMLPWNHFP	240
DB	181	PDTIAKFLFTSGSTKLPKAVPTTQRMLCANQQLQTFVPFGSEPPVLVDWMLPWNHFP	240
QY	241	SHNIGIVLYNGGYVDDSKPTAOGFAETLRNLSEISPTAYLVTPKGMELVGALESD	300
DB	241	SHNIGIVLYNGGYVDDSKPTAOGFAETLRNLSEISPTAYLVTPKGMELVGALESD	300
QY	301	LRRFPARMKLFFFAAGLSOGIWDRLDVAEHCGERIRMAAGLMTETAPSCFTTGP	360
DB	301	LRRFPARMKLFFFAAGLSOGIWDRLDVAEHCGERIRMAAGLMTETAPSCFTTGP	360

This Page Blank (uspto)

```

Db      499 GCGTAAACCTCCTGTCGATGCTTGTGGGATCAAGTATGCTATGTT----- 549
QY      93 ProTyrCysProValSerProAlaTyrSerLeu----- 104
Db      550 -----ACGTACTCTTACTCTTATGCCAGGTAATAAGCCAC 568
QY      105 -----SerGlnAspLeuAlaLysLeuArgHisIleValGlyLeuLeuGlnProGlyLeu 122
Db      589 AAGACAGTCATATGTTGTTAGAGGATCCAGATCGTCTGGAGAGATTTATGCAATATA 648
QY      123 ValPheAlaAlaAspAlaPro-----PheGln-ArgAlaIle 135
Db      649 CTGTCAATTCGAAAGTGAAGCCCTTGTGTGATAACCTGAGTCTTTCATGCAATGC 708
QY      135 eGlnThrIleLeuProAspAspAlaProAlaIlePheThrArg-----GlyGlu----- 151
Db      709 GGAATCATTTTCTTACAGACGACGCTCCAAATTTGGACTCTTCTCGGGGGGAAATC 768
QY      152 -----LeuAlaGlyArgArgThrValSerPheAspSerLeuLeuGlnProGly 168
Db      769 GTGTTGTTTACAGCGGGTAGGACACACAGCTCTATGTTACAGAAATTAAGAATT 828
QY      168 yGlyIleGlyAlaAspAsnAlaPheAlaIleThrGly----- 180
Db      829 TCGACAAAGAAACGTGCAAAATTTGCCAAGATCTATGATTCGGAAATGATGATGA 888
QY      181 -----ProAspThrIleAlaLysPheLeuPheThrSerGlySerThrLysLeuP 197
Db      889 ATATCATTCATTCACATGATATAGCCACAAATTATGATATACAGAGGAAACACAGAAATCC 948
QY      197 ouLysAlaValProThrThrGlnArgMetLeuCysAlaAsnGlnMetLeuLeuGln 217
Db      949 AAAAGGTGTTATTTCTACACATCAGAAATTTGTTACACAGATAAGAACTTGTCCGATT 1008
QY      217 rPheProValPheGlyGluGluProProValLeuValAspTyrLeuProTyrAsnHis 237
Db      1009 TGTGCT-----GCCGAAGCTGGGGAAGATTTCTGAGTATGTTCCATCATGCGATCG 1062
QY      237 rPheGlyGlySer-----HisAsnIleGlyIle----- 246
Db      1063 TTATGAACGGGCTGTGAATTACTCATATTACATGAGTGAAGCAAAAGTATACGTC 1122
QY      246 ----- 246
Db      1123 TATAGATTCTTAAGATGATCTCAAGGTTATCAACCACTATCTTCACTTC 1182
QY      247 -----ValLeuTyrAsnGlyGlyThrTyrTyrLeuAspAspGlyLysProTh 262
Db      1183 TTTAGTATATGAGACACTCTACACGCGGATTCAAAAGCAAAATTTCTGCAAGCTCCCTCG 1242
QY      262 rAlaGlnGlyPheAlaGluThrLeuArgHisLeuSerGluIleSerProThrAlaTyrIle 282
Db      1243 TCGTAAATTTTGGCACTTACATTGATCAAAAGTCAC-----CTGCAATATAC 1290
QY      282 uThrValProLysGlyTyrGluGluLeuValGlyAlaLeuGluArgSer----- 299
Db      1291 GGAATGAAAGAGTTATGAGGCTT-----TGTTTGACAAAATAATCAAAAGCTCC 1344
QY      300 -----ThrLeuArgGluArgPhePheAlaArgMetLysLeuPhePhePheAl 315
Db      1345 AATGTATATGTTTCTGTTGATGATGTTGTGGCGAGAGTAGTTCATTTTTC----- 1399
QY      315 aAlaAlaGlyLeuSerGlnGlyIleThrAspArgLeuAspArgValAlaGluGlnHis 335
Db      1400 -----CTATGGCCA-----TTGCATATGTTGCTGAAAAGCTTGT 1434
QY      335 sGlyGluArgIleArgMetMetAlaGlyLeu----- 345
Db      1435 ACAAGAAAATTCGTTGCTGCTATTCGATTAACAAGCGCTGTTAGTGAGAGTGTAG 1494
QY      345 ----- 345

```

```

Db      1495 TTACCTATGATGTTGACAAAGTTTTTTGAAGCCATCGGTGATGATGACAAATGGGTA 1554
QY      346 -GlyMetThrGluThrAlaProSerCysThrPheThrGlnProLeuSerMetAlaGly 365
Db      1555 TGGTTGACAGAAACCTCACCGGTTCTCTGGCGGAGGCTTAGGTGTAACGTTCTGG 1614
QY      365 yTyrIleGlyLeuProAlaProGlyCysGluValLysLeuVal----- 379
Db      1615 CTGATGGGCACTCATTAATAAGATACAGAGTTCAAAATTTGATGATCATGAGACTGGTAC 1674
QY      380 -----ProValAspGlyLysLeuGluGlyArgPheHisGlyProHisValMetSer 397
Db      1675 TGTTCCTTCCACCTGTTCAAGGGCATGTGCAAAAGTCAGAGGCCACCGGTATGAAAG 1734
QY      397 yTyrTyrAspAlaProGluGlnAsnAlaGlnAlaPheAspGluGlyTyrTyrCys 417
Db      1735 TTTCTACAAAGATCCACTGGCCACCAAGAGGTTATAGACATGATGATGATGATCAATAC 1794
QY      417 rGlyAspAlaIleLysLeuAlaAspProAlaAspProGlnLysGly----- 432
Db      1795 TGGAGATATGGGTGATTAATCT-----CTCAGCACTCAACAGAGCGAGTCG 1842
QY      433 -----LeuMetPheAspGlyArgIleAlaGluAspPheLysLeuSer 447
Db      1843 TAGCTGTGAGGTGTCATGCTTCTTGAAGCGCGTGCAGGAAGT-ATCGTCTTTCAC 1901
QY      447 rGlyValPheValSerValGlyProLeuArgThrArgAlaValLeuGluGlyLys 467
Db      1902 AGGT-----GAAATGTGGAACCATTTGAGATGTAAGAAGCGCCATGAAAGCAATT 1955
QY      467 rValLeuAspValValAlaAlaProAspArgLysCysLeuGlyLeuLeuValPhe 487
Db      1956 GATTCAACAAATATGATGTTATGACAGAGATCAACGCCCTTGAGACTATGTTATCC 2015
QY      487 oArgLeuLeuAspCysArgAlaLeuSerGlyLeuGlyLysGluAla----- 502
Db      2016 A-----AACAAAGAACGACAGAGAGCAGCAAGCAAAATTTCACTGTGAC 2066
QY      503 -SerAspAlaGluValLeuAlaSerGluProValArgAlaTyrPheAlaAspTyrLeu 522
Db      2067 TTCTGAAGTCATGAACCTTAGCAAGAGACGATTAACAGCATGCTTATGAAGACTAG 2126
QY      522 sArgLeuAsnArgLysAlaThrGlyAsnAlaSerArgIleMetTyrValGlyLeuLeu 542
Db      2127 GAAATGAGCGTCACATGCTCGTTCCAAAGTTGACCAAGTTCTG-----ATCGTGA 2177
QY      542 pThrProProSerIleAspLysGlyGluValThrAspLysGlySerIleAsnGlnArg 562
Db      2178 TGAACCATTCACGATAGCAACGCTTAAATGACACCGCAATGAATAAAGACGGAGCAA 2237
QY      562 aValLeuGlnTyrArgSerAlaLysValAspAlaLeuTyrArg 576
Db      2238 GGTGTGATCAATACAGAAATGAATAGAAAGACTCTTACAG 2280

```

Search completed: February 8, 2005, 02:57:24
 Job time : 4080 secs

```

QY 324 ---TTPAAspArgLeuAspArgValAlaGlu-----GlnHis-----CysGlyGluArgIle 339
Db 1345 TTCTATGGAGCGGCGCCCATGATGAGGAGACACACCTTCTTCTGTGGTCTCAACATC 1404
QY 340 ArgMetMetAlaGlyLeuGlyMetTrpGluThrAlaProSerCysThrPheThrThrGly 359
Db 1405 CGCTTGATGCGGGGCTACGGCTCAGTGAACCTCA---GGCCCCAATTCAATGTCAGT 1461
QY 360 ProLeuSerMetAlaGlyTyr---IleGlyLeuProAlaProGlyCysGluValIleLeu 378
Db 1462 CCTTCAACACTACCGGCTGTACAGCTCAGGCAAGTTGTGCTCGCGGTGGAAGCTG 1521
QY 379 ValProValAspGlyIleLeuGluGlyArg-----PheHisGlyProHisValMetSer 396
Db 1522 GTGAACACAGACGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1581
QY 397 GlyTyrTrpArgAlaProGluGluAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 416
Db 1582 GGCTACCTGAACATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1641
QY 417 SerGlyAspAlaIleIleLeuAlaAspProAlaAspProGluIleGlyLeuMetPheAsp 436
Db 1642 AGCGGAGGAGTGGCGCGCTG-----GAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1689
QY 437 GlyArgIleAlaGluAspPheIleLeuSerSerGlyValPheValSerValGlyProLeu 456
Db 1690 GGGCGGCTCAAGAAATTAATCATCAGACGCTGTGGGAGAGATGCGCCCTGTGCCATC 1749
QY 457 ArgThrArgAlaValLeuGluGlySerTyrValLeuAspValValAlaAlaAlaPro 476
Db 1750 GAGGAGGCGGCTGAAGTGAAG---CTGCCCATCATCAGCAAGCCCATGCTCATGGGAGC 1806
QY 477 AspArgGluCysLeuGlyLeuValPhe-----ProArgLeuLeu 490
Db 1807 CAGAGGAAGTCTCTGCTCATGCTGCTCACCCTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1866
QY 491 Asp-----CysArgAlaLeuSerGlyLeu 498
Db 1867 GACCAAGCATGATTAATGACTGAACAGCTGTGAGCTTGTCCAGGAGGAGGAGGAGGAGGAG 1926
QY 499 GlyIleGluAlaSerAspAlaGluValLeuAlaSerGluProValArgAlaTrpPheAla 518
Db 1927 GCCACACAGTGTCCAGATCATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1986
QY 519 AspTrpLeuIleArgLeuAspArgGluAlaThrGlyAspAlaSerArgIleMetTrpVal 538
Db 1987 GAGGAGGATCCGAGGAGGATGAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2046
QY 539 GlyLeuLeuAspThrProProSerIleAspIleGlyValThrAspIleSerIle 558
Db 2047 GCCATTCTGAGAGAGACTTCTTCATTTCCGCTGAGAGAGTGGTCCCAAGTAAACTG 2106
QY 559 AsnGlnArgAlaValLeuGlnTrpArgSerAlaValAspAlaLeuTyrArgGlyGlu 578
Db 2107 AAACGGCTACAGCTTTTGGAGAGATCAAAAGATCATGACTCTCTTAC-----CAA 2160
QY 579 AspGlnSerMet 582
Db 2161 GAGCAAAAATG 2172

```

```

RESULT 15
LOCUS 2366 bp mRNA linear HTC 06-FEB-2004
DEFINITION Arabidopsis thaliana Full-length cDNA Complete sequence from clone
GSLTPGH6ZG05 of Hormone Treated Callus of strain col-0 of
Arabidopsis thaliana (thale cress).
ACCESSION BX824843
VERSION 1 GI:42466280
KEYWORDS HTC; GSLT; cDNA;
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```

```

REFERENCE
AUTHORS
1 (bases 1 to 2366)
Castelli,V., Aubry,J.M., Jallion,O., Wincker,P., Clepet,C.,
Menard,M., Cruaud,C., Quetier,F., Scapellato,C., Schachter,V.,
Temple,G., Caboche,M., Weissenbach,J. and Salanoubat,M.
Whole Genome Sequence Comparisons and 'Full-Length' cDNA Sequences:
A Combined Approach to Evaluate and Improve Arabidopsis Genome
Annotation
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 2366)
Genoscope.
Direct Submission
Submitted (18-NOV-2003) Genoscope (E-mail : seqref@genoscope.cns.fr
BP 191 91006 Evry cedex - FRANCE)
Web : www.genoscope.cns.fr
COMMENT
The sequences are based on single pass reads.
Life Technologies (a division of Invitrogen) members carried out
full-length libraries construction : Temple G.
Genoscope members carried out sequencing and annotation : Castelli
V., Aubry J.M., Jallion O., Wincker P., Menard M., Cruaud C.,
Schachter V., Weissenbach J., Salanoubat M.
URGV INRA : Clepet C., Caboche M.
Annotation is based on the June 2003 version of the Arabidopsis
genome released by MIPS (Munich Information center for Protein
Sequences). 5 prime and 3 prime are assembled with Phrap.
http://www.genoscope.cns.fr/externe/sequences/Banque_Projet_EF/Full
Length
http://www.genoscope.cns.fr/cgi-bin/ggb/ggb?source=Arabidopsis.
FEATURES
source
1..2366
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/strain="Col-0"
/db_xref="taxon:3702"
/collection="GSLTPGH6ZG05"
/issue_type="Hormone Treated Callus"
/plasmid="pCMVSPORT_6"
1..2366
/gene="At3g23790"
ORIGIN
Alignment Scores:
Pred. No.: 7,36e-12 Length: 2366
Score: 226.00 Matches: 151
Percent Similarity: 35.24% Conservative: 108
Best Local Similarity: 20.54% Mismatches: 263
Query Match: 7.33% Indels: 214
Gaps: 29
DB:
US-09-750-986D-30 (1-589) x CNS0A7C (1-2366)
QY 2 ArgSerLeuGluAlaLeuLeuProPheProGlyArgIleLeuGluArgLeuGluIleTrp 21
Db 238 AAGAGGATTAAGCGCGGTTCGCATTC-----TTAGAAAGCTTA----- 276
QY 22 AlaIleThrArgProGluGlnThrCysValAlaAlaAlaAlaAlaAlaAlaAlaAla 41
Db 277 -----TCATTGCCAAGGAG-----GCTGCTTGGAGCTTAATGATGAG 318
QY 42 -----ArgIleSerTyrAla 46
Db 319 TCTGTTCTCGATATTGAGATCATCTGTGAGAAATACGATGACAGATGCGGAGTGA 378
QY 47 GluMetPheHisAsn-----ValArgAlaIleAlaGlnSerLeuLeu 60
Db 379 GATCCGATATCAGACCGGCTTCTACATTCACGATACAGACAGTGGAGAAATCTTG 438
QY 61 ProTyr-----GlyLeuSerAlaGluArgProLeuLeuIleVal 73
Db 439 GACTTTGTTGAGGCTTACAGAGTCTGTGAAGTAAAGAGAGAGAGAGATGCACTTTT 498
QY 74 SerGlyAspAspLeuGluIleLeuGluAlaPheGly---AlaMetTyrAlaGlyIle 92

```

QY 338 gileargmetcmetalaGlyLeu----- 345
::: |||
Db 1659 TCMAACATGGGCGGCTGCCGAGTGCCAGCTGATTTCTCTCAGCGMAACAT 1718
QY 346 -----GlymetrhglutrhralaProSerCyethrPhehrthrGl 359
|||::: |||
Db 1719 CCTATCATGAGTTGATGACGAGCTTGGAG-----TGACGGGACTACACAC 1766
QY 359 yProleusermetalaGlyTYr-----IleGlyLeuProalaProGl 373
::: |||
Db 1767 AGTGTCACACCTGCAGGCTTACCGGATCTGACGCGTGGGAAAGCGCTTCCAAAACCA 1826
QY 373 yCySGluValLyLeuValaProValaSpGlyLyLeuGluGlyYarPhehisglProhI 393
::: |||
Db 1827 CACAAAGGTGAAAAAGAAACAAAGATGGGATCGGGAACCTTGATCTGCGGCGCCCA 1886
QY 393 sValmetSerGlyTYrTYrPargAlaProGluGlnaAmAlaGlnAlaPheAspGluGlu 413
::: |||
Db 1887 CATATTCAATGGGCTACCTCAGACACAGACACAGACAGAGAGGTGACACCCACCG 1946
QY 413 yTYrTYrCySerGlyAspAlaIleLyLeuValaAspProAlaAspProGlnLyglYle 433
|||::: |||
Db 1947 CTGGCTACACACGACGACTTGGGCTTCTTG-----GATTTGACAAAGTTCT 1994
QY 433 yMetPheAspGlyYargIleAlaGluAspPheLyLeuSerSerGlyValaPheValSerVa 453
|||::: |||
Db 1995 CATATGCATGGGACAAATACATTCATTCATCAAGCTGAGCTCGGGGAAATGTCACACC 2054
QY 453 lGlyProLeu-----ArgThrArgAlaValleuGluGlySerTyValle 469
|||::: |||
Db 2055 GATGCCCATCGAGAGCGGGTAAAGAGCGGAATCCCATC-----GTCCG 2099
QY 469 uAspValValaAlaAlaProAspArgGluCyLeuGlyLeuLeuValaPheProArgle 489
::: |||
Db 2100 CATATGCATGGGCTGGGCGGACGAGCGCCGCTACCTGTGCGCCCTCTC-----AC 2150
QY 489 uLeuAspCyArgAlaLeuSerSerGlyLyGluGluValaSerAspAlaGluValleuAl 509
|||::: |||
Db 2151 ATTGAAGTGTCAATTAATCAAGACGATGAGAGCGTCCGAGTAC-----CTGAC 2201
QY 509 aSerGluProVal-----ArgAlaTrpPheAlaAspTr 520
|||::: |||
Db 2202 CAGTGAAGTGTCTGCTTGTGCGAGGCTTCGACCGACGCTGCTGTTCGACACG 2261
QY 520 pLeu-----LysArgLeuAs 525
|||::: |||
Db 2262 CCTATATGACCGTGAACCTTTGGTCAACGAGTTCAATTAACAGGCGATCGAGATGAA 2321
QY 525 nArgGluAla-----ThrGlyAsnAlaSerArgIleMetTrpValGlyLeuLeuAspTrp 544
|||::: |||
Db 2322 TTCGAGGCGGCCCAAAATGGGCGCAAGATCAATTAAGTGGTC---ATTCTAACAACGA 2378
QY 544 oProSerIleAspLyGlyGluValaThrAspLyGlySerIleAsnGlnArgAlaValle 564
|||::: |||
Db 2379 TTTCTCTGTGTGTGTGGGAGCTTGACCCATGACCAAGATGAACAGATCCATTGTGG 2438
QY 564 uGlnTrpArgSerAlaLyValaAspAlaLeuTYr 575
::: |||
Db 2439 CAAGATATACCAAGAGAAATCCAGAAACTCTAT 2472

RESULT 14
AY418305 2175 bp DNA linear GSS 12-DEC-2003
LOCUS AY418305
DEFINITION Homo sapiens HCM6513 gene, VIRUTAL TRANSCRIPT, partial sequence,
genomic survey sequence.
ACCESSION AY418305
VERSION AY418305.1 GI:39774265
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 2175)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Shtnsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 2175)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Shtnsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT These sequences were made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source
1. 2175
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
<1..>2175
/locus_tag="HCM6513"
ORIGIN
Alignment Scores:
Pred. No.: 6.47e-12 Length: 2175
Score: 226.00 Matches: 102
Percent Similarity: 38.15% Conservative: 75
Best Local Similarity: 21.98% Mismatches: 213
Query Match: 7.33% Indels: 74
Gaps: 14
US-09-750-986D-30 (1-589) x AY418305 (1-2175)
QY 181 ProAspThrIleAlaLyAspPheLeuPhehrSerGlySerThrLyLeuProLyAlaVal 200
|||::: |||
Db 817 CCCAACAGTGTGTGTGCTGCTACGTCTACCTCCGGGACCACTGGAAACCCAGGGCGTG 876
QY 201 ProThrThrGlnArgMetLeuCyAlaAsnGlnMetLeuGlnThrPheProVal 220
|||::: |||
Db 877 ATGCTGATCAAGACATATACGTCGACGCGACGCGTACGCGACGCGGTGATC 936
QY 221 PheGlyGluGlu-----ProProValLeuValaAspTrpLeuProTrpAsnHisThrPhe 238
|||::: |||
Db 937 CGGCGCGCAGAAAGTCCAGACGAGAGGTGTAGTACCTGCTGCTCCCTCAAGCATATTGCC 996
QY 239 GLyGlySerHisAsnIleGlyIleValleuTYrAsnGlyGlyTYrTYrLeuAspAsp 258
|||::: |||
Db 997 GCCCAATCTACGACCTGTGACAGGCAATCCAGTGGGGGCCCAAGTTTCTTGGCGAA 1056
QY 259 GlyLyPProThrAlaGlnGlyPheAlaGluThrLeuArgAsnLeuSerGluLeSerPro 278
|||::: |||
Db 1057 CCCGAGCGCCCTGAAGGGAGAGCTGTGAACAACGCTCGG-----GAGGTGAGGCC 1107
QY 279 ThrAlaTyLeuThrValaProLySgLYrTrpGluGluLeuValaGlyAlaLeuGluAsp 298
|||::: |||
Db 1108 ACATCACACATGTGGGCTGCCCGGTATGAGAAAGATCATGTGAGCGCATCCAG--GAG 1164
QY 299 SerThrLeuArgGluTrpPhePheAlaArgMetLyLeuPhePheAlaAlaGly 318
|||::: |||
Db 1165 GTGGCGGCTCAAGTGTGCTTCAATCCGCGAAGATCTGTGTGGCATGTGGTGACC 1224
QY 319 LeuSerGlnGlyLe----- 323
|||::: |||
Db 1225 TTGAGACAGAAACCTCACTGCGCCCGGACGACGACCTGAAGCCCTTCACAAACAGACTGGCA 1284
QY 323 ----- 323
Db 1285 GATTACTGTGTCTAGCCAAAGTTCCGACGACCTGGGATTTGGCAAGTGTCAAAAGAAC 1344

QY	107	AspLeuAlaIaIyLeuArgHisIeIeValGlyLeuLeuGlnProGlyLeuValPheAlaIa	126
Db	584	GAA-----GCCATGACGTACTGATATCAACAAGCTGAACCTCTCTGGATTTTGTGAC	637
QY	127	AspAlaIaProPheGlnArgAlaIeGlyThrIleLeuProAspValProIaIa---	145
Db	638	AAGCAGAGAAAGGCCAAACCTTTATTAGAGGGGTGAAGAAATAAGTTAAATCAAGCCTT	697
QY	146	-----IlePheThrArgGlyGlyLeuAla	153
Db	698	AAATCATAGTTGTCATGATGCTTACGCGCAGTGAACCTGGTGAACGAGGCCAGAGGTGT	757
QY	154	GlyIArgArgThrValSerPheAspSerLeuLeuGlnProGlyGlyIleGlyAlaIaAsp	173
Db	758	GGGGTGAAGATCACACGATGAAGCGC--ATGAGAGA-CCTGGGAAGAGCCAAACAGACG	813
QY	174	AsnAlaPheAlaIaIaThrGlyProAspThrIleAlaIaIySpheLeuPheThrSerGlySer	193
Db	814	GAAGGCCAAGGCTCCAGCA--CCTGAATGATCTTGAGTAATTTGTTTTCACAAGTGAACCT	871
QY	194	ThrIySLeuProIySAlaIaValProThrThrGlnArgMetLeuCyAlaAsnGlnIleMet	213
Db	872	ACGGGCAACCCCAAGAGACAAATGCTCACTACCGAAACATATGAGGAGATTTGTCACCT	931
QY	214	LeuLeuGlnThrPheProValPheGlyGlyGluProPro-----ValIleValAspTrp	231
Db	932	TTTGTGAAGAACAGAAATACAGTACATCTTCCACATGATATCTTGATATCTTTC	991
QY	232	LeuProTrpAsnHisIaThrPheGlyGly-----SerHisAsnIle	244
Db	992	TTGCCTCTCGCCCATATGTTGAGAGAGTTGAGAGTGTGAATGCTGTATGATGAGACT	1051
QY	245	GlyIleValLeuTyraSngIyGlyThrTyTrIleuAspAspGlyLysProThrAlaGln	264
Db	1052	AAATTCGAGATTTTCCAGAGAGATTCAGCTGCTCAGAT-----	1093
QY	265	GlyPheAlaGluThrLeuArgAsnLeuSerGluIleSerProThrAlaTyIleuThrVal	284
Db	1094	-----GACTCTCAAGGCTTCAACCACTGCTCCCGCTGGTT	1132
QY	285	Pro-----LysGlyTrpGluIleuValGlyAlaLeuGluArgAspSerThr	300
Db	1133	CCAGACTGCTGAACCGGATGTTTACCGAATTTTCGA-----CAAGCAAAACCAACG	1186
QY	301	LeuArgGlu-----ArgPhePheAlaArgMetLysLeuPhePheAlaIaIaIa	317
Db	1187	CTGAAGCATGGCTCTTGACCTTTCCTCCAAAGAGAA-----GAGACA	1231
QY	318	GlyLeuSerGlnGlyIle-----TrpAspArgLeu--AspArgVal	330
Db	1232	GAGCTTGCAGACGGGATCATCAGAAACAACAGCCTGTGGAGACGGGCTGATCTTCCACAA	1291
QY	331	AlaIaGlnHisCySgIyGlyArgIleArgMetMetAla-----	343
Db	1292	GTAACAGTGAACCTCGGCGGAAGAGTCCGGCTGATGTGACAGAGCCGCCCGGTGCT	1351
QY	344	-----GlyLeuGly	346
Db	1352	GCCACTGTGTCAGCTTCTCTAGACAGACCCCTGGGCTCTCAGTTTATTGAAGAAATGAAATAC	1411
QY	347	MetThrGluThrAlaProSerCysThrPheThrThrGlyProLeuSerMetAlaGlyLys	366
Db	1412	CAGACAGAGTGCATCGCCGGGGTGGCTTGACCATGCTCGAGACTGACGACGCGACAT	1471
QY	367	IleGlyLeuProAlaProGlyCySgIeValIySLeuValProVal-----	381
Db	1472	GTTGGGGCCCGATGCCGTGCAATTGTATAAACTTTGTGATGTGGAAGAATGAAATAC	1531
QY	382	-----AspGlyLysLeuGlnGlyLysArgPheHisGlyProHisValMetSerGlyLys	398
Db	1532	ATGCTCGACGAGGGGAGAGGTGTGTGTGAAGGCGCAAAATGATTTTCAAGGGCTAC	1591

QY	399	TrpArgAlaProGluGlnAsnAlaGlnAlaPheAspGluGluGlyTrpTrpCysSerGly	418
Db	1592	TTGAGAGACCCAGCGAAAACAGCAGAAAGCTTTGGACAAAGCGCTGGTTACACACAGGG	1651
QY	419	AspAlaIleIysIleuAlaAspProAlaAspProGlnIysGlyLeuMetPheAspGlyArg	438
Db	1652	GACATTGGAAAAATGGTTA-----CCAAATGGCACCTTGAATAATTATGCACGG	1699
QY	439	IleAlaGluAspPheIysIleuSerSerGlyValPheValSerValGlyProLeuArg--	457
Db	1700	AAAAAGACATATTTAAGCTGGCACAAGAGAAATACATAGCCCTCGAAAGATTGAAAT	1759
QY	458	-----ThrArgAlaValLeuGluGlyGlySer-----	466
Db	1760	ATCTACATCGCAAGTGAAGCTGTGTCAGGTGTTTGTCCACGAGAAAGCTTCGACGCA	1819
QY	467	TyrValLeuAspValValAlaAlaProAspArgGluCysLeuGlyLeuLeuValPhe	486
Db	1820	TTTCTCATGTGAATGTGGTA-----CCAGATTTTAGACATTA-----	1858
QY	487	ProArgLeuLeuAspCysArgAlaLeuSerGlyLeuGlyIysGluAlaSerAspAlaGlu	506
Db	1859	-----TGTTCTCGGGCCCAAAAGAGAGGATTTGAAGGTCGTTTGAAGAA	1903
QY	507	ValLeuAlaSerGluProValArgAlaTrpPheAlaAspTrpLeuIysArgLeuAsnArg	526
Db	1904	CTGGCGCAAAATATGAGTCTCAAAAAGACTATCTCGCAAGATATGCTGAGCTTTGGAG	1963
QY	527	GluAla---ThrGlyAsnAlaSerArgIleMetTrpValGlyLeuLeuAspThrProPro	545
Db	1964	GATTCTGGTCTGAAACCATTTGGAACAGAGCTCAAAAGCATTCACATTGCAACCTGAAATTATT	2023
QY	546	SerIleAspIysGlyGluValThrAspIysGlySerIleAsnGlnArgAlaValLeuGln	565
Db	2024	TCTATCGACATGCGCTTTCGATCTCAACAATGAAGCGAAAAGCCAGAGCTCGGAAC	2083
QY	566	TrpArgSerAlaIysValAspAlaLeuTyr	575
Db	2084	TATTTACAGTTCGACGATGATGATGACTCTAT	2113

	RESULT 13
LOCUS AKO28078	
DEFINITION Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:1700026H02 product:hypothetical AMP-dependent synthetase and ligase containing protein, full insert sequence.	
ACCESSION AKO28078	
VERSION AKO28078.1 GI:26080715	
KEYWORDS HTc, CAP trapper.	
SOURCE Mus musculus (house mouse)	
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
REFERENCE 1 Carninci,P. and Hayashizaki,Y. High-efficiency full-length cDNA cloning <i>Meth. Enzymol.</i> 303, 19-44 (1999) JOURNAL MEDLINE 99279253 PUBLISHED 10349636	
REFERENCE 2 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Komo,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cdnas to prepare full-length cDNA libraries for rapid discovery of new genes <i>Genome Res.</i> 10 (10), 1617-1630 (2000) JOURNAL MEDLINE 20499374 PUBLISHED 11042159	
REFERENCE 3 Shibata,K., Itoh,M., Aizawa,K., Nagasaka,S., Sasaki,N., Carninci,P., Komo,H., Akiyama,J., Nishi,K., Kitamura,T., Tashiro,H., Itoh,M., Suni.N., Ishii.Y., Nakamura.S., Hazema.M., Nishire.T., Harada.A., Yamamoto.R., Matsumoto.H., Sakaguchi.S., Ikegami,T., Kashiwagi.K., Fujiwake.S., Inoue,K., Togawa,Y., Izawa.M., Ohara,E., Wataniki,M.	


```

Db      1543 ATATCAAGCTCGTGTGATGTGAGAACTGAACCTAGCGCTCGAAAGGAGAGAGAGA 1602
Qy      387  lYarPheHISglYProHISvalMeSerGlYTrTParGAlaProGluAlaSnlaG 407
Db      1603 TATGTGTGAGAGAGAACCAATGTGTTCAAAGGCTTCTTGAAGATCCGAGACGAGCAAG 1662
Qy      407  lAlaPheAspGluGluGlyTrYTrYTrYCySerGlYAspAlaIleYLeuAlaAspPro 427
Db      1663 AGGCGCTGAGACAGCATGGCTGCTTCACTGAGACATCGGAAATGCGT----- 1715
Qy      427  lAspProGluInlySGlyLeuMetPheAspGlyArGlleAlaGluAspPheYLeuSers 447
Db      1716 -----CCGGCAGGAACCTTAAATATATGATCGGAAACATATATTTAACTTGCTC 1770
Qy      447  erGlYValPheValSerValGlyPro-----LeuArgThrArgAlaVal- 461
Db      1771 AGGGAAGATATGTGTGACCCGAGAGATTGAGAACATCTACATCCGAGCGCACTGTGG 1830
Qy      462  -----LeuGluGlyGlySer-----TyrValLeuAspValValAlaAla 475
Db      1831 CGCAATCTATGTCCATGGGAGACGTTAAAGGCTTTTGGTAGGCAATTGTGTG---- 1866
Qy      475  lAProAspArgGluCySLeuGlyLeuLeuValPheProAlGLeuLeuAspCyArGlAl 495
Db      1887 --CCTAACCTCGAA-----GTTATGCCCTCTCGGCGCCAGAGAGA---- 1925
Qy      495  euserGlYLeuGlyGlySGluAlaSerAspAlaGluValLeuAlaSerGluProValArg 515
Db      1926 -----GGATTTGAGAGACATATGCAGATCTCTGCACAAATAGAGATCTGAAGA 1974
Qy      515  lATrPheAlaAspTrPLeuYsArgLeuAsnArgGluAla---ThrGlyAsnAlaSer 534
Db      1975 AAGCCATTTTGGAAGATATGTGTGAGTTAGGAAAGAAAGTGAATCCATCTTTTGAGC 2034
Qy      534  rglLeuMetTrPValISglYLeuLeuAspThrProProSerIleAspYlSGlyGluValThra 554
Db      2035 AGGTTAAAGCATTCATCATCTTCGACATGTTCTCATGTTCAAAATGGCTTCTGCAC 2094
Qy      554  sPrYsGlySerIleAsnArgAlaValLeuGlnTrPArgSerAlaYLeuValAspAla 574
Db      2095 CAACACTAAAGCTTAAGAGACCTGAGCTGAGAGAGTACTTCAAAAACAAATAGAGAGC 2154
Qy      574  euTYr 575
Db      2155 TTATAC 2159

RESULT 11
AY402444
LOCUS      2097 bp      DNA      linear      GSS 15-DEC-2003
DEFINITION Homo sapiens FAC12 gene, VIRTUAL TRANSCRIPT, partial sequence,
ACCESSION  AY402444
VERSION     AY402444.1  GI:39758430
KEYWORDS   GSS.
SOURCE      Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 2097)
            Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarimal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Slnsky,J.J.,
            Adams,M.D. and Cargill,M.
            Inferring nonneutral evolution from human-chimp-mouse orthologous
            gene citos
JOURNAL    Science 302 (5652), 1960-1963 (2003)
PUBMED     14671302
AUTHORS    Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarimal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Slnsky,J.J.,
            Adams,M.D. and Cargill,M.

```

```

TITLE      Direct Submission
JOURNAL    Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
            Rockville, MD 20850, USA
COMMENT    This sequence was made by sequencing genomic exons and ordering
            them based on alignment.
FEATURES   Location/Qualifiers
            source
              1..2097
              /organism="Homo sapiens"
              /mol_type="genomic DNA"
              /db_xref="taxon:9606"
              <1..>2097
              /gene="FAC12"
              /locus_tag="HOM126"
ORIGIN
Alignment Scores:
Pred. No.:      8.18e-14      Length:      2097
Score:          244.00      Matches:      139
Percent Similarity: 36.89%      Conservative: 96
Best Local Similarity: 21.82%      Mismatches: 261
Query Match:    7.92%      Indels:      143
DB:             Gaps:      22

US-09-750-986D-30 (1-589) x AY402444 (1-2097)
Qy      30  CyValAlaAlaArgAlaAlaenGlyGluTrPArgArgIleSerYTrAlaGluMetPhe 49
Db      322 TGTATTAGGCTCGGAAACCGACCAACCTATGATGCTTTCATATTAACAGTTGCA 381
Qy      50  HisAsnValArgAlaIleAlaGlnSerLeuLeuProTYrGlyLeu-----SerAlaGlu 67
Db      382 GAATTGTGCGAGTGCATAGGCTCAGCATGATCAGAAAGGCTTCAAGACTGCCAGAT 441
Qy      68  ArgProLeuLeuIleValSerGlyAsnAspLeuGluHisIleuGlnIleuAlaPheGlyAla 87
Db      442 CAGTTATTTGGCATCTTTGCTCAAAATPAGACCTGAGTGGGTATTTTGACAAAGATGC 501
Qy      88  Met---TyrAlaGlyIleProTYrCySProValSerProAlaYrSerLeuLeuSerGln 106
Db      502 TTTCGTTATTCAGATGCTG-----ATCGTTCCACTTATGATACCTTGGAAT 549
Qy      107  AspLeuAlaYleuLeuArgHisIleValISglYLeuLeuGlnProGlyLeuValPheAlaAla 126
Db      550 GAA-----GCCATCAGATACATAGTCAACAAAGCTCTCTGCTGTTTGTGAC 603
Qy      127  AspAlaAlaProPheGlnArgAlaIleGluThrIleLeuProAspArgValProAla--- 145
Db      604 AAGCCAGAGAGGCCAAACTCTTATTAGAGGCTGTGAATAATAGTTAATACGAGCCTT 663
Qy      146  -----IlePheThrArgGlyGluLeuAla 153
Db      664 AAATCATAGTTGTCTCATGATGCTTACGGCAGTGAAGACTGTGAGACGAGCCAGAGGTGT 723
Qy      154  GlyArgArgThrValSerPheAspSerLeuLeuGluGlnProGlyGlyIleGluAlaAsp 173
Db      724 GGGGTGAAGTCAACCGCATGAAGGCG---ATGGAAGA--CTTGGGAAGAGCCAAAGACG 779
Qy      174  AsnAlaPheAlaAlaThrGlyProAspThrIleAlaYrPheLeuPheThrSerGlySer 193
Db      780 GAAGCCCAAGCTCCGCA--CCTGAAGATCTTGCAGTATTGTTTCAACATGGAACCT 837
Qy      194  ThrYLeuProYsAlaValProThrThrGlnArgMetLeuCySAlaAsnGlnMet 213
Db      838 ACAGGCAACCCCAAGAGACATGTCATCCAGCAAAACATAGTACGCAATTGTCAGCT 897
Qy      214  LeuLeuGlnThrPheProValPheGlyGluGluProPro-----ValLeuValAspTrp 231
Db      898 TTGTGAAGCAACAGAGATACAGTCAATCTTGCCAGATGATACTTGATATCTTTC 957
Qy      232  LeuProTrpAsnHisPheThrPheGlyGly-----SerHisAsnIle 244
Db      958 TTGCTCTGCGCCCATATGTGTTGAGAGAGTTGAGAGTGTGTAAGCTGTGATCGTAGAGCT 1017

```



```

Qy 67 -----GluArgProLeuLeuLeuValSerGlyAsnAspLeu 78
Db 596 CAGTTATTATGTTGTTTGACAAAATCGCCAGAGGATCATTTG-GGAGCTGGGCTG 654
Qy 79 GluHLeuGlnLeuAlaPheGlyAlaMetTyrAlaGly-IlePro---TyrCysProVa 97
Db 655 CTACACATA-----TTCCATGGTGGTGTCCCTCATATACACCTT 696
Qy 97 IserProAla-----TyrSerLeuSerGlnAspLeuAlaLys----- 110
Db 697 GGGCCCTGGGGCTATCCGCTACATCATATACAGCGACATACAGCAGCTGATTGGGA 756
Qy 111 -----LeuArgHisIleValGlyLeuLeuGlnProGlyLe 122
Db 757 CAAACCTCAGAAAGCTGTGCTTCTGTAGAGCATGTGAGAGAGAGAGACCTCAGGCTT 816
Qy 122 uValPheAlaAlaAspAlaAlaProPheGlnArgAlaIleGluThrIleLeuProAspAs 142
Db 817 CAAGCTGATCATCTTCATGAGACCATTCAGAAAGAGCCCTGAAA----- 859
Qy 142 pValProAlaIlePheThrArgGlyLeuLeuAlaGlyArgThrValSerPheAspSe 162
Db 860 -----GAGAGAGGCGAGAAAGTCCGGGTGGTATTAACTTCATCGACGC 903
Qy 162 rLeuLeuGlnGlnProGlyGlyIleGluAlaAspAsnAlaPheAlaAlaThrGlyProAs 182
Db 904 CGTG-----GAGGACGTGGCCAGAGAAATCACAGGCTCCTGTGCCCCGAGCCTGA 957
Qy 182 pThrIleAlaLysPheLeuPheThrSerGlySerThrLysLeuProLysAlaValProth 202
Db 958 TCACCTCTCATTTGTGTGTTTTCAGAGCGGACGAGAGAACCCAAAGAGTGCATGCT 1017
Qy 202 rThrGlnArgMetLeuCyAlaAsnGlnGlnMetLeuGlnThrPhePro-----Va 220
Db 1018 CACCCTAGGGAACGTGGCTGATTTCTCAGGCTTTCTGAAGTACAGAGAAAGTAT 1077
Qy 220 lPheGlyGlnGlnProProValLeuValAspTrpLeuProTrpAsnIleThrPheGly 240
Db 1078 CTTTCGAGACAGAGACGATGTGCTCATTCCTGCTGCTGCTGCTGCTGCTGCTGCTG 1132
Qy 240 ySerHisAsnIleGlyIleValLeuTyr---AsnGlyGlyThrTyrTyrLeuAspAspG 259
Db 1133 -GAGAGAGTATTCAGATCTGTGCTTATGTGCCAGAGGCGCTGT----- 1177
Qy 259 yLysProThrAlaGlnGlyPheAlaGlu---ThrLeuArgAsnLeuSerGlu----- 275
Db 1178 -----GGCTTCTTCCAGGAGATATCCGCTTCTCATATACATGA 1221
Qy 276 ---IleSerProThrAlaTyrLeuThrValPro----- 285
Db 1222 GGCTATATGCCCCACATCTTCCCTGTGGTCCACAGACGTGTGACCGGATGTACGACA 1281
Qy 286 -----LysGlyTTP----- 288
Db 1282 GATCTTCACGACGAGCAAAACACACATTAAAGCGCTGCTCTCGAGTGTTCAGCAAGCG 1341
Qy 289 -----GluGlnLeuValGlyAlaLeuGlnIuArgAspSerThrLeuArgGlnuArgP 305
Db 1342 TAAGCAAGCCGAGAGTCCGAGAGTGAATCATCAGAAATGATATGTCTGGGATGAACCTT 1401
Qy 305 ePheAlaArg-----MetLysLeuPhePhePheAlaAlaAla 317
Db 1402 CTTTAAATAGATTACAGCCAGCTGTGTGGGTGTGGCGGATGATTGTACTGGAGCAGC 1461
Qy 317 aGlyLeuSer-GlnGlyIleTrpAspArgLeuAspArgValAlaGluGlnHisCysGlyG 337
Db 1462 CCCAGCATACCAACAGTCTCTGGAT-----TTCCTCGGGCG 1497
Qy 337 lu-----ArgIleArgMetMetAlaGlyLeuGlyMetThrGlnuThrAlaProSerCys 355
Db 1498 AGCTCTAGGGGTGCGAGATTATGA-GGTTATGCGCAAACTGAGTGACACGCTGGATGTA 1556
Qy 355 hPheThrThrGlyProLeuSerMetAlaGlyTyrIleGlyLeuProAlaProGlyCysG 375

```

```

Db 1557 CTTTCACCACTCTCGGCGACTGTGACCTTCAGGGGACGATAGGGGCGCCACTTCCCTGCATTC 1616
Qy 375 luValLysLeuValProValAsp-----GlyLysLeuGlnug 387
Db 1617 ATATCAAGGCTCGTGTAGTGTGTGGAACTGAATCTACTGGGCTGCAAAAGAGAGGAGAGA 1676
Qy 387 lYArgPheHisGlyProHisValMetSerGlyTyrTrpArgAlaProGlnIuAsnAlaG 407
Db 1677 TATGTGTAGAGAGACCAAAATGTGTTCAAAAGGCTACTTGAAGAATTCAGACAGAGCAAGG 1736
Qy 407 lAlaPheAspGlnGlnGlyTyrTyrCysSerGlyAspAlaIleLysLeuAlaAspProA 427
Db 1737 AGGCGCTGACAGCGATGTGCTGCTTACACCTGGAACATCGGAAATATGGCTG----- 1789
Qy 427 lAspProGlnLysGlyLeuMetPheAspGlyYArgIleAlaGluAspPheLysLeuSer 447
Db 1790 -----CCGGCAGGAACCTTTAAATTTATTTATGATGAAAAAGCATATATTAACTTGCTC 1844
Qy 447 eArgIValPheValSerValGlyPro-----LeuArgThrArgAlaVal- 461
Db 1845 AGGAGAAATATATGTGACCCGAGAAAGATTGAGAAACATTAATCCGAGCGCAACTGTGG 1904
Qy 462 -----LeuGlnGlyGlySer-----TyrValLeuAspValValValAlaAla 475
Db 1905 CGCAATCTATGTCCATGTGGGAGCAGCTTAAAGCCTTTTGTGTAGCATTTGTGTG----- 1960
Qy 475 lAspAspArgGlnGlyCysLeuGlyLeuLeuValPheProArgLeuLeuAspCysArgAlaL 495
Db 1961 --CTGACCTGAA-----GTTATGCCCTCTCTGGGCCAGAGAGA----- 1999
Qy 495 euserGlyLeuGlyGlyGlnAlaSerAspAlaGluValLeuAlaSerGlnProValArgA 515
Db 2000 -----GGAATTTGAAGGAACATATGACATCTCTGCACAAAATAGATCTGAGA 2048
Qy 515 lATrPheAlaAspTrpLeuLysValArgLeuAsnArgGluAla---ThrGlyAsnAlaSerA 534
Db 2049 AAGCCATTTTGGAGATATGTGTCAGGTGTAGAAAGAAAGTGCATCTCATTTCTTTGAGC 2108
Qy 534 rglMetTrpValGlyLeuLeuAspThrProProSerIleAspLysGlyGluValThra 554
Db 2109 AGGTTAAAGCATTCATCATTCATCTTGCACATGTCTTCAAGTTCAAAATGGCTTGACAC 2168
Qy 554 ePlySerIleSerIleAsnGlnArgAlaValLeuGlnInTrpArgSerAlaLysValAspAlaL 574
Db 2169 CAACACTAAAGGTAAAGACCTGAGCTGAGAGATGTAATTTCAAAAACAAATGAAGAGC 2228
Qy 574 eUTyr 575
Db 2229 TTATAC 2233

```

RESULT 10
 BC026161
 LOCUS
 DEFINITION
 Homo sapiens acyl-CoA synthetase long-chain family member 6, mRNA
 (CDNA clone IMAGE:4830357), containing frame-shift errors.
 ACCESSION
 BC026161
 VERSION
 BC026161.1 GI:20072873
 KEYWORDS
 HTC.
 SOURCE
 Homo sapiens (human)
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 2518)
 REFERENCE
 AUTHORS
 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
 Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
 Altschul, S.F., Zeeberg, B., Buetow, K.H., Scheffer, C.F., Bhat, N.K.,
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, D., Hsieh, F.,
 Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
 Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
 Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
 Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
 Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J.,

```

Db      703 TTTATTAAGCAAGAGAGTCGATTCATCGCTTCCACAGATATGTCGATATCTTTC
Qy      232 LeupProTyrAsnIleThrPheGlyGly-----SerHisanIle 244
Db      763 TTGGCTCTCGCCCATATGTTTGAGACCGGTGTAGAGTGTATTAATGCTGTGATGAGCT 822
Qy      245 GYLLLeuLeuTyrAsnGlyGlyThrTyrTyrLeuAspAspGlyLysProThrAlaGln 264
Db      823 AAGATAGGATTTTCCAGAGATATCAAGCTGCTTATGAC-----864
Qy      265 GlyPheAlaGluThrLeuArgAsnLeuSerGluIleSerProThrAlaTyrIleuThrVal 284
Db      865 -----GACCTCAAGTGCTTCAAGCCCAACATCTTCCCTGCGGTT 903
Qy      285 ProLysGlyTyrPheGluLeuValGlyAlaLeuGluArgSerThrLeuArgGlyArg 304
Db      904 CCCAGG-----CTGCTGAACCGG 921
Qy      305 PhePheAlaArg-----MetLysLeuPhePhePheAlaAlaAla-----317
Db      922 ATGTTGACAGATGGAACAAACAACTTCTTGAAGCATGAGCTGTTGACTTTCCTC 961
Qy      318 -----GlyLeuSerGln-GlyIle-----TyrAspArg 326
Db      982 CAAAGGAAAGAGCGCGAGCTTGCAGTGCATCGTCAGAAACAGACGCTGGGATTA 1041
Qy      326 GLeu--AspArgValAlaGluGlnHisCysGlyGlyArgGlyLeuMetMet-----342
Db      1042 ACTCATCTTCCACAAAGATACAGTCGAGCCTCGGTGGGAAAGTCGGCTGATGATCAAG 1101
Qy      342 -----342
Db      1102 AGCAGCCCGGTGCTGCCACAGTCTGACGTTTCTGAGACAGCCCTGCGCTCAAGTT 1161
Qy      343 ----AlaGlyLeuGlyMetThrGluThrAlaProSerCysThrPheThrThrGlyProLe 361
Db      1162 CATATGAAGCTACGACAGACGAGTGCATCGTGTGTGCTGCTGAGCTGCGCGGAGA 1221
Qy      361 uSerMetAlaGlyTyrTyrIleGlyLeuProAlaProGlyCysGluValLysLeuValProVa 381
Db      1222 CTGACGCGAGCGCATGTTGAGCCCCCATGCTTCCAAATTATGTAAGCTTGTGATG 1281
Qy      381 lAsp-----GlyLysLeuGluGlyArgPheHisGlyProH 393
Db      1282 GGAAGAAATGATTAATTCCTGCGCATCAAGGCGAGGCTGAGTGTGTGAAAGGGCGAAA 1341
Qy      393 eValMetSerGlyTyrTyrPArgAlaProGluGlnAsnAlaGlnAlaPheAspGluGlu 413
Db      1342 TGTGTTCAAAAGGTAATTGAAAGACCAGACAGACAGCTGAGCCCTGATTAAGATG 1401
Qy      413 yTyrTyrCysSerGlyAspAlaIleLysLeuAlaAspProAlaAspProGlnLysGlyLe 433
Db      1402 CTGCTTACACACGCGGAGCATTTGAAAATGCGCTG-----CCAATGCGACCTT 1449
Qy      433 uNecPheAspGlyArgIleAlaGluAspPheLysLeuSerSerGlyValPheValSerVa 453
Db      1450 GAAAGATTATCGACAGAAAAACATATTAACTAGCCCAAGAGAGATACATAGACACC 1509
Qy      453 lGlyPro-----LeuArgThrArgAlaVal-----LeuGluG 464
Db      1510 AGAAAAGATTGAATATCTACTGCGGAGTGAAGCGTGGCCAGATGTTTGTCCACGG 1569
Qy      464 yGlySer-----TyrValLeuAspValValValAlaAlaProAspArgLysGly 481
Db      1570 AGAAAGCTTCAAGCGCTTCTCTACAGATGTGGTA-----CCGACGCTTATAGAGCCT 1623
Qy      481 uGlyLeuValPheProArgLeuLeuAspCysArgAlaLeuSerGlyLeuGlyLysG 501
Db      1624 A-----CGTCTCGGCGACAGAAAGAGGCTTCAAGGG-----1657
Qy      501 uLAserAspAlaGluValLeuAlaSerGluProValArgAlaTyrPheAlaAspTyrle 521

```

```

Db      1658 ----TCCTTGAAGAACTGTGCGAGAAACAGATATCATTAAGCTATCTCGAGCACTT 1713
Qy      521 uLysArgLeuAsnArgGluAlaThrGlyAsnAlaSerAlaGlyMetTyrValGlyLeu 541
Db      1714 GTTGAACCTTGGAGAGAAAGCC---GGTCTGAAGCATTTGAACAGCTCAAAAGCAATTGC 1770
Qy      541 uAspThrProPro-----SerIleAspLysGlyGluValThrAspLysGlySerIleAs 559
Db      1771 TGTGACCCCGGAATTAATTTCTATTTCACAAAGCGCTTCTGATCTCAACACTGAAGCGAA 1830
Qy      559 nGlnArgAlaValLeuGlnTyrArgSerAlaLysValAspAlaLeuTyr 575
Db      1831 GAGGCGAGAGTACGAACTATTTCAGTCGACAGATAGTAAGTCTTAC 1879

RESULT 9
LOCUS      CR606980      2548 bp      mRNA      linear      HTC 21-JUL-2004
DEFINITION      Full-length cDNA clone CS0DJ003YJ20 of T cells (Jurkat cell line)
               Cot 10-normalized of Homo sapiens (human).
ACCESSION      CR606980.1 GI:50487787
KEYWORDS      HTC; CNSL1_cDNA.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
REFERENCE      1 (bases 1 to 2548)
AUTHORS      Li, W.-B., Gruber, C., Jessee, J. and Polyaes, D.
TITLE      Full-length cDNA libraries and normalization
JOURNAL      Unpublished
CONTACT      Feng Liang Email : fliang@lifetech.com URL :
               http://fulllength.invitrogen.com/InvitrogenCorporation 1600
               Faraday Avenue
               2 (bases 1 to 2548)
REFERENCE      Genoscope.
AUTHORS      Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
JOURNAL      BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
               - Web : www.genoscope.cns.fr)
COMMENT      1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
               end enriched, double-strand cDNA was digested with Not I and cloned
               into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
               was normalized. Library was constructed by Life Technologies, a
               division of Invitrogen.
FEATURES             Location/Qualifiers
     source           1..2548
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DJ003YJ20"
                     /issue_type="T cells (Jurkat cell line) Cot
                     10-normalized"
                     /plasmid="pCMVSPORT_6"
ORIGIN
Alignment Scores:
Pred. No.:      1 286-14      Length:      2548
Score:          253.00      Matches:      148
Percent Similarity: 38.10%      Conservative: 97
Best Local Similarity: 23.02%      Mismatches: 246
Query Match:    8.21%      Indels:      154
DB:             3      Gaps:      27

US-09-750-986D-30 (1-589) x CR606980 (1-2548)
Qy      30 CysValAlaAlaArgAlaAlaAsnGlyLutPArgArgIleSerTyrAlaGluMetPhe 49
Db      476 TGTCTTGTTTCAAGAGAGCTTACAGAGCTTACAGTGTGCTTACACAGAGAGTGGCC 535
Qy      50 HisAsnValArgAlaAlaAlaGlnSerLeuLeuProTyrGlyLysLeuSerAla-----66
Db      536 GACAGGCGTGAATTTCTGTGGGTCGCGAGCTTCCACAGACAAATTGTAAGCATGCACTGAT 595

```

Db 1382 ACAGAGCAGACCCGGGTCTGCCACAGTGTGACGTTTCTGAGGACGAGCGCTGGCTGC 1441
Qy 343 -----AlaGlyLeuGlyMetThrGluThrAlaProSerCysThrPheThrGly 359
Db 1442 CAGTTTATGAGGCTACGACGACGAGCGATGACGCTGCTGCTCCCTGACCTTGGCCC 1501
Qy 360 ProLeuSerMetAlaGlyTyrTyrIleGlyLeuProAlaProGlyCysGluValIlyLeuVal 379
Db 1502 GGAGACTGACGCGACGCGCATGTTGGAGCCCCCATGCTTGCATTAATGTAAGCTTGTG 1561
Qy 380 ProValAsp-----GlyLeuGluGluGlyAArgPheHisGly 391
Db 1562 GATGTGGAGAAATGAATTAATTAATCTGGATCCAGGCGGAGGTGATGTGTGAAAGGG 1621
Qy 392 ProHisValMetSerGlyTyrTrpArgAlaProGluGlnHisAlaGlnAlaPheAspGlu 411
Db 1622 GCAATGTGTCTCAAGGCTTACCTTGAAGAGCCAGCAGAAACAGCTGAAGCCCTGGATAA 1681
Qy 412 GluGlyTyrTyrCysSerGlyAspAlaIleLeuLeuAlaAspProAlaAspProGlnIly 431
Db 1682 GATGCTGGGTACACACGCGGGGACATTTGAAATGGCTG-----CCAAATGGC 1729
Qy 432 GlyLeuMetPheAspGlyTyrIleAlaGluAspPheIlyLeuSerSerGlyValPheVal 451
Db 1730 ACCTTGAAGATTATGACAGGAAACACATATTAACTACCCAGAGAGATACATA 1789
Qy 452 SerValGlyPro-----LeuArgThrArgAlaVal-----Leu 462
Db 1790 GCACCGAAGAAAGTTGAAATATCTACTCTGCGAGTGAAGCCCTGGCCCGACGCTTTGTC 1849
Qy 463 GluGlyGlySer-----TyrValLeuAspValValAlaAlaPheAspArgGlu 479
Db 1850 CACGGGAAAGCTTTCAGAGCCCTTCTCATAGCAGTTGTGGTA-----CCGACGTTGAG 1903
Qy 480 CysLeuGlyLeuLeuValPheProArgLeuLeuAspCysArgAlaLeuSerGlyLeuGly 499
Db 1904 AGCCTA-----CGTCCCTGGGACAGAGAGAGGCTTTACAAAGG-----1942
Qy 500 LysGluAlaSerAspAlaGluValLeuAlaSerGluProValArgAlaTrpPheAlaAsp 519
Db 1943 -----TCTTCGAAAGACTGTGCAGAACAGATATCAATTAAGCTATCTCGAC 1993
Qy 520 TrpLeuLysArgLeuAsnArgGluAlaThrGlyAsnAlaSerArgIleMetTrpValGly 539
Db 1994 GACTGTGTGAACCTTGGAGAGAGGCC---GGTGAAGCATTGTAACAGTCAAGG 2050
Qy 540 LeuLeuAspThrProPro-----SerIleAspIlyGlyValIlyThrAspIlySer 557
Db 2051 ATTGCTGTCCACCCGGAATTAATTTCTATTGACACGCGCTTGTGACTCCACACCTGAAG 2110
Qy 558 IleAsnGlnArgAlaValLeuGlnTrpArgSerAlaIlyValAspAlaLeuTyr 575
Db 2111 GCGAAGAGCCAGAGCTACGGAATATTTCAGGTCCGACATGATGAACCTGTAC 2164

RESULT 8
AY402446 1897 bp DNA linear GSS 15-DEC-2003
LOCUS Mus musculus FACL2 gene, VIRTUHL TRANSCRIPT, partial sequence.
DEFINITION genomic survey sequence.
ACCESSION AY402446
VERSION AY402446.1 GI:39758432
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1897)
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejaritwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE
Inferring nomenclature evolution from human-chimp-mouse orthologous
gene titos

JOURNAL Science 302 (5652) , 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 1897)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejaritwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source
1. 1897
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
<1. >1897
/gene="FACL2"
/locus_tag="HCM1226"
gene
ORIGIN
Alignment Scores:
Pred. No.: 8.11e-15 Length: 1897
Score: 253.00 Matches: 154
Percent Similarity: 36.89% Conservative: 81
Best Local Similarity: 24.18% Mismatches: 259
Query Match: 8.21% Indels: 143
DB: 9 Gaps: 27
US-09-750-986D-30 (1-589) x AY402446 (1-1897)
Qy 30 CysValAlaAlaArgAlaIleAlaAsnGlyGluTrpArgArgIleSerTyrAlaGluMetPhe 49
Db 127 TGTATTGATGTTCTGGAAGCCAAACAGCCCTATGATGATTTCTTCAAAAGAGGTGCA 186
Qy 50 HisAsnValArgAlaIleAlaGlnSerLeuLeuProTyrGlyLeuSer-----AlaGlu 67
Db 187 GAATCGCTGAGTGAATGATGAGCTCCGGGCTGATGCAGAAAGGTTCAAGCTTCTCGAG 246
Qy 68 ArgProLeuLeuIleValSerGlyAsnAspLeuGluHisLeuGlnLeuAlaPheGlyAla 87
Db 247 CAGTTTCATGCGGCTCTTCTTCAAAACAGACCGAGTGGGTGATCTCGACGAAGATGC 306
Qy 88 Met-----TyrAlaGlyIlyProTyrCysProValSerProAlaTyrSerLeuLeuSerGln 106
Db 307 TTCTCTTACTCATGTGT-----GTGCTCCGCTCTATACACCTTGGAGCT 354
Qy 107 AspLeuAlaLysLeuArgHisIleValGlyLeuLeuGlnProGlyLeuValPheAlaIle 126
Db 355 GAC-----GCCATCACCTACATATGTAACAAAGCTGAACCTCTGTGATTTTGTCTGAC 408
Qy 127 AspAlaIleProPheGlnArgAlaIleGluThrIleLeuProAspAspValProAla--- 145
Db 409 AAGCCAGAAAAGACCAACTTATTAGAAAGGTGTAAGAAACAAAGTTAACACCATGCCCTT 468
Qy 146 -----IlePheThrArgGlyGluLeuAla 153
Db 469 AAAATATATGTCATCTACGACCTCTACGCGAGTATCTGTTGAGACGAGCAAGAAAGTGT 528
Qy 154 GlyArgArgThrValSerPheAspSerLeuLeuGluGlnProGlyGlyIleGluAlaAsp 173
Db 529 GGGGTGGAATATCATGCTCAAGACT---CTGAGAGACCTTGGAGAGAGTGAACAGAGTG 585
Qy 174 AsnAlaPheAlaIleThrGlyProAspThrIleAlaLysPheLeuPheThrSerGlySer 193
Db 586 AAGCCC---AAGCTTCAGAACCCGGAAGATCTTGCAATATTTGTTCAACAGTGAACCT 642
Qy 194 ThrIlyLeuProLys---AlaValProThrThrGlnArgMetLeuCysAlaAsnGlnGln 212
Db 643 ACAGGCAACCCCAAGAGAGCATGATCACTACCAAAACATTTTAAACAGCTCTCAGGT 702
Qy 213 MetLeuLeuGlnThrPheProValPhe---GlyGluGluProProValIleuValaAspTrp 231

QY	287	GLYTRPGIUGLIEUVALGLVALLEUENLURGRASERTHRLIEARGLIAPHEPHE	306
Db	1193	ATTTAIGCTGCAATTACAAATGCTGTGAAGAGCTCTNNNNNTTGAAGAAAGTTGTAT	1255
QY	306	-----	306
Db	1253	CATACGTGATCAATGCGACAGCATGCAATGATTAATGTGAGAAACCCATCACTTATG	1312
QY	307	-----	312
Db	1313	TGGGACAAGTTGATTTAAACAAATTAAAGCTTGAAGCTTGTGGACGATAGAGGCTTATG	1372
QY	313	PHEPHEALALALAGLYLEUSERGLNGLYLETTPRASPARGLEUASPARGLVALALAGLU	332
Db	1373	ACTTCAGGTGCTCTTCACATGTGTCCACAGATGTCATGAA-----NNNNNA	1417
QY	333	GLNIHISCYS---GLYGLUARGILEARGMETALAGLYLEUGLYMETTHRGILUTHRALA	351
Db	1418	AGAAATATGCTTCGGTGAA-----CTTCTGAAGGATATGGAATGACAGAGACA---	1465
QY	352	PROSERCYSINRPHETHRTHR-----GLYPROLEUSERMETALAGLYTYRILEGLY	368
Db	1466	---TCCTGTGCATCAGTACAGTACATGAAATTTGGTGACAGATCAAT---GGAACATGTGGA	1519
QY	369	LEUPROALAPROGLCYCYSGULVALLYSEUVALPROVALAP-----	382
Db	1520	TCCTCAAAATCCTTGTGTGAGGTTAAACTTGTGACGTCGCCAGAAATGAATATATCTTGG	1579
QY	383	-----GLYLYSEUENLUGLYATSPHEHLEGLYPROHISVALMET	395
Db	1580	GAGATCAACGGTATCCTCGTGAGAGATTTGTGTAG-----GACCTATATATATTC	1633
QY	396	SERGLYTYTPARGALAPROGLGLNANALAGLINALAPHEAPGLUGLYTYRTRY	415
Db	1634	TGGGGTTACTATTAAGATGAGTCCAAACAGAGAGCTATTGATGAGATGGTGGCTTG	1693
QY	416	CYSSERGLYAPRALALIELYLEUALASPRALAPROGLNLYGLYLEUMETHE	435
Db	1694	CATCTGAGACATA-----GGTCGTGGCTG	1720
QY	436	ASPGLY-----ARGILEALAGLUSPHELYLEUSERGLY	448
Db	1721	CCTGAGAGGCGCTTAAAGATCATGATGAGNNNNGAACATTTCAAGTTAGCTCAAGGA	1788
QY	449	VALPHEVALSERVALGLYPROLEUARG-----THARG	459
Db	1781	GAATACATAGCTCCAGAGAGATGGAATAATGTCATGCAATGCAATGTCATTTGCCCAA	1844
QY	460	ALVALLEUENLUGLYGLYSETRY-----VALLEUASPRVALVALALALAPRO	476
Db	1841	TGTTTTTATATGTGGCGACAGTTTCAACTTCTTGATGTGATTTGTGACAGTTGAAACCG	1900
QY	477	ASP-----ARGGLCYSEUENLYLEU	483
Db	1901	GATGTTATGAAGGCTTGAGGCTGCATCTGAAGGATCCAGTGCAGAGACTTGAGCAGCTT	1960
QY	484	LEUVALPHEPROARG-----LEUENLAPCYARGALALEUSERGLYLEU	498
Db	1961	TGTGTGACCTTGAAGCAAGAGCTGCAGACTCTGGCCGAC-----ATGATATCTGTT	2011
QY	499	GLYLYSGILUALA-----SERAPRALAGLUALLEUALASERGLU	511
Db	2012	GGGAGAGGACACAGCTGAGAGGTTTGGAAATTGTCTAAAGCTGTACTCTGTGGCTGAG	2071
QY	512	PRO-----VAL	513
Db	2072	CCGTTTCACTTGATATGCGCTTCTCACCCCAACTTCAAGTCAAGAGCCCGCAGCC	2131
QY	514	ARGALATRPHEALASPTRLLEUARYGLEUBENARGILUALATHRGILYASNALASER	533
Db	2132	AAAGCATATCTTTGCA-----AAAGAAATCTCAAGACATGATTTTCA	2170
QY	534	ARGILEWETTRPVALGLYLEUENLAPTHRPROPROSERILLEAPLYSGLYGLUVALTHR	553

Db	2171	CGGTGCTAGAGACAGACTTACGATGCTCAAGTTGTAACCTTGCGGTATCAGATCATG	2230
Oy	554	Aplysglyserilleanglinargala-----ValleugIntP-----	566
Db	2231	GAGAAAGGTCTATCATACAGCCGTCACAGAGATTGATTCGCAATTGTTATTCACCT	2290
Oy	567	-----ArgserAlaIyValAlaApAlLeuTyArgIy---GluApArgInserMetIeu	583
Db	2291	CACCAAAATATGATGTAATAATCTNNNNNACTGATACGGGAATATACGGATCATGCTC	2350
Oy	584	Arg 584	
Db	2351	CGG 2353	
RESULT 7			
AK004897		2848 bp	linear
LOCUS			HTC 03-APR-2004
DEFINITION		Mus musculus adult male liver cDNA, RIKEN full-length enriched	
		library, clone:1300004819 product:fatty acid Coenzyme A ligase,	
		long chain 2, full insert sequence.	
ACCESSION			
VERSION		AK004897.1	GI:12836428
KEYWORDS		HTC; CAP trapper.	
SOURCE		Mus musculus (house mouse)	
ORGANISM		Mus musculus	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
TITLE		Carninci, P. and Hayashizaki, Y.	
JOURNAL		High-efficiency full-length cDNA cloning	
MEDLINE		Mech. Enzymol. 303, 19-44 (1999)	
PUBMED		99279253	
REFERENCE		10349636	
AUTHORS		2	
		Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,	
TITLE		Itch, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.	
JOURNAL		Normalization and subtraction of cap-trapper-selected cDNAs to	
MEDLINE		prepare full-length cDNA libraries for rapid discovery of new genes	
PUBMED		Genome Res. 10 (10), 1617-1630 (2000)	
REFERENCE		20499374	
AUTHORS		11042159	
		3	
		Shibata, K., Itch, M., Aizawa, K., Nagaoke, S., Sasaki, N., Carninci, P.,	
		Komoto, H., Akiyama, J., Nishi, K., Kitunai, T., Tashiro, H., Itch, M.,	
		Sun, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,	
		Yamanote, R., Matsunoto, H., Sekaguchi, S., Ikegami, T., Kashiwagi, K.,	
		Pujiwako, S., Inoue, K., Togawa, Y., Irawa, M., Ohara, E., Matshiki, M.,	
		Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunura, S., Kawai, J.,	
		Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.	
		RIKEN integrated sequence analysis (RISA) system--384-format	
		sequencing pipeline with 384 multichipillary sequencer	
		Genome Res. 10 (11), 1757-1771 (2000)	
JOURNAL		20530913	
MEDLINE		11076861	
PUBMED		4	
REFERENCE		The RIKEN Genome Exploration Research Group Phase II Team and the	
AUTHORS		FANTOM Consortium.	
TITLE		Functional annotation of a full-length mouse cDNA collection	
JOURNAL		Nature 409, 685-690 (2001)	
REFERENCE		5	
AUTHORS		The FANTOM Consortium and the RIKEN Genome Exploration Research	
TITLE		Group Phase I & II Team.	
JOURNAL		Analysis of the mouse transcriptome based on functional annotation	
REFERENCE		of 60,770 full-length cDNAs	
AUTHORS		Nature 420, 563-573 (2002)	
TITLE		6 (bases 1 to 2848)	
JOURNAL		Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H.,	
REFERENCE		Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y.,	
AUTHORS		Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hizumoto, K.,	
TITLE		Hirakawa, T., Hori, F., Imotani, K., Ishii, Y., Itch, M., Izawa, M.,	
JOURNAL		Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kono, H., Kouda, M.,	
REFERENCE		Koyu, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K.,	
AUTHORS			

Db 1733 ACTGCGGGCGCGAAGCGTCCCGGTGCTATCGAGAGCGCGTGAAGATGAG--- 1789
Qy 465 GlySerThrValLeuAspValValAlaAla-----ProAspArgGluCys 480
Db 1790 -----CTGCCCATCATCAGCATGCTCTGATTCGGGGACAGAGAGATTC 1837
Qy 481 LeuGly-----LeuLeuValPheProArgLeuLeuAsp----- 491
Db 1838 CTGTCCATGCTGTCTTACCTTGAAAGTCAAGCTGAGCCAGAGAGCTTGAGCTTACAGAC 1897
Qy 492 -----CyArgAlaLeuSerGlyLeuGlyGlyValAla 502
Db 1898 AGCGTCAGCCAGAACGCTGTGAGTCTGCGAGCGGTGGGACGACAGACGACCGTC 1957
Qy 503 SerAspAlaGluValLeuAlaSerGluProValArgAlaTrpPheAlaAspTrpLeuLys 522
Db 1958 TCCGAGATCGTGGGCGAGAGAGAGAGCGCTGTACAGAGCCATTCAGAGAGATTCAG 2017
Qy 523 ArgLeuAsnArgGluValThrGlyAsnAlaSerArgIleMetTrpValGlyLeuLeuAsp 542
Db 2018 AGGGTGAACCGAAGCCGACGCGCCCTTACCATTCAGAAAGTGGCCATTTCTCAA 2077
Qy 543 ThrProSerIleAspGlyGlyValThrAspLysGlySerIleAsnGlnArgAla 562
Db 2078 CGTACCTTCTCATTGAGGTGAGAGAGCTGGAGCCCATGAATGAAGAGCGCTCAGC 2137
Qy 563 ValLeuGlnTrpArgSerAlaLysValAspAlaLeuTrpArgGlyLeuAspGln 580
Db 2138 GTTCTGAGAGAGTAAAGATATCATCATCTTTTATCATAGAGCAAAAGCAG 2191

RESULT 6
AY109410 2505 bp mRNA linear HTC 17-OCT-2002
LOCUS Zea mays CL741.1 mRNA sequence.
DEFINITION AY109410
ACCESSION AY109410.1 GI:21213124
VERSION
KEYWORDS
SOURCE
ORGANISM
Zea mays
Zea mays
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
Maize Mapping Project/Dupont Consensus Sequences for Design of
Overgo Probes
Unpublished (2002)
2 (bases 1 to 2505)
Coe,E.H.
Direct Submission
Submitted (25-APR-2002) Maize Mapping Project, University of
Missouri, Columbia, MO 65211, USA
If you are interested in getting corresponding physical clones,
these are publicly available from ZmDB and may be found by BLAST
searching at MSL, maizegap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
maize cDNA sequences is either Virginia Walbot, Stanford or Pat
Schubler, Iowa State, then clones may be requested from ZmDB:
www.zmdb.iastate.edu.
Location/Qualifiers
1..2505
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="taxon:4577"
/clone_id="Maize Mapping Project/Dupont Consensus
Library"
/note="this sequence is part of a project of EST
assemblies resulting from the application of public
contigs to seed Dupont contigs; this resource was
assembled by Dupont as part of a collaboration for the
overgo addressing of BACs in conjunction with the Maize

ORIGIN Mapping Project"
Alignment Scores: 7.91e-16 Length: 2505
Pred. No.: 264.50 Matches: 154
Score: 36.62 Conservative: 110
Percent Similarity: 21.368 Mismatches: 220
Best Local Similarity: 8.588 Indels: 237
Query Match: 3 Gaps: 37
US-09-750-986d-30 (1-589) x AY109410 (1-2505)
Qy 23 LysThrArgProGluGlnThrCysValAlaAlaArgAlaAlaAsnGlyGluTrpArg 42
Db 425 AGAATCGCCCAACGAGGAGATT-----GAGAGTCAAGTGG 463
Qy 43 IleSerThrAlaGluMetPheIleAsnValArgAlaIleAlaGlnSerLeuProTrp 62
Db 464 ATGACGTATGAGAAAGCTAGACGACAGACAGAACTGCAATAGTTCTGTATTATTCAT 523
Qy 63 GlyLeuSer-----AlaGluArgProLeuIleVal 73
Db 524 GGAATATCTGAAGTGCACGAAATGCTATATTATTAACACGAGCGGATGATCAT 583
Qy 74 SerGlyAsnAspLeuGlnIleGluGlnLeuAlaPheGlyValMetTrpAlaGlyIlePro 93
Db 584 -----TTGACCATGCTTGTGGCATCTCA-----TATGATCTGTG--- 622
Qy 94 TyrCysProValSerProAlaTyrSerLeuLeuSerGlnAspLeuAlaLysLeuArgHis 113
Db 623 -----CGGCTTATGATCTCTGCGCCAGAT-----GCAGTTCAGTTT 661
Qy 114 IleValGlyLeuLeuGlnProGlyLeuValPheAlaAlaAspAlaAlaProPheGlnArg 133
Db 662 ATATGTAACCAACGCAACGATGAGATTAATCTGTGTG-----CCCCAA 706
Qy 134 AlaIleGluThrIleLeuPro----- 140
Db 707 ACTCTAAGCCCTATTTTAAGCTTTATTACTCAAAATCCGCTGTCTGTTNNNNNNNN 766
Qy 141 -----AspAspValProAlaIlePheThrArgGlyGlyLeuAlaGlyArg 155
Db 767 NNNNNNNNNNNNTTCAATATATGATCATCTGACGACGACAACTGTAATGCC----- 817
Qy 156 ArgThrValSerPheAspSerLeuLeuGlnProGlyGlyIleGluAlaAspAla 175
Db 818 AAATATATGACGTACACAGGCTACCAATGAG-----GAAAGATGAGTCTCAAACT 871
Qy 176 PheAlaAlaThrGlyProAspThrIleAlaLysPheLeuPheThrSerGlySerThrLys 195
Db 872 TTTCGTCACCAAAACCTGAAGATATAGCCATATCTGCTACACGATGACATCTAGGC 931
Qy 196 LeuProLysAlaValProThrThrGlnArgMetLeuCysAlaAsn-----GlnGln 212
Db 932 ACACAAAGGAGGAGCTGATCTTCTCATGAGAACTTAATGCAATAGAGAGGCTCAAGT 991
Qy 213 MetLeuLeuGlnThrPheProValPheGlyGluGlnProProValAlaLeuAlaTrpLeu 232
Db 992 TTAGGATTAAGTTTACCC-----TCGATGTGTACATCTGTTCTA 1036
Qy 233 ProTrpAsnIleThrPheGlyGlySerHisAsnIleGlyIleLeuValLeuTrpAsnGly--- 251
Db 1037 CTTTGGCTCACAATCATGAGAGAGCTAACACAGAGTGCATCTTCACTATGTTGCC 1096
Qy 252 ---GlyThrTyrTrpLeuAspArgGlyLysProThrAlaGlnGlyPheAlaGluThrLeu 270
Db 1097 ATTGATTTCAACAGAGGATTAATCTGAAG-----CTGATG 1132
Qy 271 ArgAsnLeuSerIleSerProThrAlaTyrLeuThrValPro-----Lys 286
Db 1133 GATGATTTGCTGTCTGAGACCAACAAATATTCGAAGGTTCCTCGGTTATATAACAA 1192

Db 243 ATGTATCGGCTGGGGCTGTCGGAACATCGCCGCTGTGCACCGGTCACATTTCTC 184
Qy 360 oLeuSerMetAlaGlyTyrIleGlyLeuProAlaProGlyCyseGluValIysLeuValPr 380
Db 183 GCGGAGCTCCCGTGAATATCGCGCTTCGGAATACCGGTCCTGAGCTGAAGCTTCGAC 124
Qy 380 oValAspGlyLysLeuGluGlyIleGlyPheHISGlyProHISValMetSerGlyTyrTrp 400
Db 123 CTCGGCGCAACAGCTGTAAAGTCGCGCTGTCGTCGCCAACGTCACGCTGGGTTACTGAA 64
Qy 400 gAlaProGluGlnAsnAlaGlnAlaPheAspGluGluGlyTyrTyrCyseSerGlyAspAl 420
Db 63 GCGCGCGAGCTGACCAAGGAGCGTTCGTGCAGAGGCGCTTTATCTCTCTCGTGAATGC 4
Qy 420 a 420
Db 3 G 3
RESULT 3
AZ933744/c 714 bp DNA linear GSS 24-APR-2001
LOCUS B1_Ba0001A14r B. japonicum BAC library Bradyrhizobium japonicum
DEFINITION genomic, genomic survey sequence.
ACCESSION AZ933744.1 GI:13775804
VERSION AZ933744.1
KEYWORDS Bradyrhizobium japonicum
SOURCE Bradyrhizobium japonicum
ORGANISM Bradyrhizobium japonicum
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Bradyrhizobiaceae; Bradyrhizobium.
REFERENCE 1 (bases 1 to 714)
AUTHORS Tomkins,J.P., Wood,T.C., Stacey,M.G., Loh,J.T., Judd,A.,
Goicoechea,J.L., Stacey,G., Sadowsky,M.J. and Wing,R.A.
A marker-dense, sequence-ready map of the Bradyrhizobium japonicum
genome Res. 11 (8), 1434-1440 (2001)
JOURNAL genome Res. 11 (8), 1434-1440 (2001)
MEDLINE 21376150
PUBMED 11483585
COMMENT Contact: Wing RA
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: twing@clemson.edu
Class: BAC ends
High quality sequence start: 23
High quality sequence stop: 664.
Location/Qualifiers
1. 714
/organism="Bradyrhizobium japonicum"
/mol_type="genomic DNA"
/strain="USDA110"
/db_xref="taxon:375"
/lab_host="E. coli"
/clone_lib="B. japonicum BAC library"
/note="Vector: pIndigo56; Site_1: HindIII"
ORIGIN
Alignment Scores: 6.33e-22 Length: 714
Pred. No.: 315.00 Matches: 77
Score: 52.23% Conservative: 40
Best Local Similarity: 34.38% Mismatches: 105
Query Match: 10.22% Indels: 3
DB: 8 Gaps: 1
US-09-750-986d-30 (1-589) x AZ933744 (1-714)
Qy 202 TbThThGlaArgMetLeuCySaIaAenGlnMetLeuLenJn-----ThrPhePro 219
Db 703 ACACCGAGGAGCATGATGACTCGAGCCAGCGCAAGCGCAAGCGCATGTCGATTCCTTG 644

Qy 220 ValPheGlyGluGluProProValIleuValAspTrpLeuProTrpAsnHISThrPheGly 239
Db 643 GACATGGCTGGACGCGATCTCGATTCGATCGAGCTGCGCTGATGCTCATCATTCGCGC 584
Qy 240 GlySerHISAsnIleGlyIleValIleuValIleuValIleuValIleuValIleuValIleu 259
Db 583 TCCAACCAACCTTTCAACCTGTGAATGCCGATCGGCGGCTCACTATATGACGCGCGC 524
Qy 260 LysProThrAlaGlnGlyPheAlaGluTrpLeuValAsnLeuSerGluIleSerProThr 279
Db 523 GAGCTTCGACCGCGTCTCTTGGAGCTGCGCTGCCAAGCTTAACAGCGGATGCCGAGC 464
Qy 280 AlaTyrLeuThrValProLysGlyTyrGluGluValGlyAlaLeuGluValAspSer 299
Db 463 GTCTATTTCAAGCTGCGCGCGCTTCACATGCTGATGCGCTGATGCTGCGTGAAGAA 404
Qy 300 ThrLeuArgGluArgPhePheAlaArgMetLysLeuPhePheAlaAlaGlyLeu 319
Db 403 GCGTTCCTCCCGCTTTCACGAGGTGAAGTTCGCTTTCACGCGGCTGAGCGCTG 344
Qy 320 SerGlnGlyIleTrpAspArgLeuAspArgValAlaGluGlnHISCyseGlyValArgGly 339
Db 343 CCGCAAGACCTGTGAACGCGCTGTAACAATTGCTCCGTCAATACCGTCGCGCGCTG 284
Qy 340 ArgMetMetAlaGlyLeuGlyMetThrGluThrAlaProSerCyThrPheThrGly 359
Db 283 CCGATGTATTCGCGCTGAGGCTGACGGAACCTCGCGCTGCGACCGCTCCATTTTC 224
Qy 360 ProLeuSerMetAlaGlyTyrIleGlyLeuProAlaProGlyCyseGluValIysLeuVal 379
Db 223 CTCGCGGAGCGCTCCGTAATATGCGCGCATACCGGTACCGAGCTGAAGCTCGTC 164
Qy 380 ProValAspGlyLysLeuGluGlyIleGlyLeuProAlaProGlyCyseGluValIysLeuVal 399
Db 163 ACCTCGGCGCCAGCTGATAGTCCGCGCGGCGCCACCGCTTCGCCAGCTTATCA 104
Qy 400 ArgAlaProGluGlnAsnAlaGlnAlaPheAspGluGlyTyrTyrCyseSerGlyAsp 419
Db 103 ACAGCTCGTG-CTGCGCAGAAAGCATAGCCGACCATGCTTTATCTCATGTGGGT 45
Qy 420 AlaIleLysLeu 423
Db 44 GCGGTAAAGCTT 33
RESULT 4
AZ935488 544 bp DNA linear GSS 24-APR-2001
LOCUS B1_Ba0003J24r B. japonicum BAC library Bradyrhizobium japonicum
DEFINITION genomic, genomic survey sequence.
ACCESSION AZ935488.1 GI:13778288
VERSION AZ935488.1
KEYWORDS Bradyrhizobium japonicum
SOURCE Bradyrhizobium japonicum
ORGANISM Bradyrhizobium japonicum
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Bradyrhizobiaceae; Bradyrhizobium.
REFERENCE 1 (bases 1 to 544)
AUTHORS Tomkins,J.P., Wood,T.C., Stacey,M.G., Loh,J.T., Judd,A.,
Goicoechea,J.L., Stacey,G., Sadowsky,M.J. and Wing,R.A.
A marker-dense, sequence-ready map of the Bradyrhizobium japonicum
genome
genome Res. 11 (8), 1434-1440 (2001)
JOURNAL genome Res. 11 (8), 1434-1440 (2001)
MEDLINE 21376150
PUBMED 11483585
COMMENT Contact: Wing RA
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: twing@clemson.edu
Class: BAC ends
High quality sequence stop: 499.

ORIGIN

```

/strain="USD110"
/db_xref="taxon:375"
/lab_host="E. coli"
/clone_lib="B. japonicum BAC library"
/note="Vector: pIndigo536; Site_1: HindIII"

```

Alignment Scores:

Pred. No.:	1,21e-42	Length:	708
Score:	514.00	Matches:	102
Percent Similarity:	58.05%	Conservative:	35
Best Local Similarity:	43.22%	Mismatches:	97
Query Match:	16.68%	Indels:	2
DB:	8	Gaps:	1

US-09-750-986d-30 (1-589) x A2933866 (1-708)

```

Qy 188 LeuPheThSerGlySerThrLyLeuProLyAlaProThrThrGlnArgMetLeu 207
   |||||
Db 708 ATTTTCCTTAAGTCTNCGCGGGGCCCAAAACCTTATCACACCCCAAGCATTTTT 649
Qy 208 CysAlaBngInMetLeuLeuGlnThrPheProValPhe-----GlyGluPro 225
   |||||
Db 648 GCCTTAGAGCAACAGGCGCAAGGCGAGNCTGCGCTTCTTAACAAGGCGCAACAT 589
Qy 226 ProValLeuValLeuProLyLeuProThrPheGlnArgMetLeu 245
   |||||
Db 588 CTGCTCATGTATGACCTGCGCTGCGCGCCACCACTTGGCGCAACCAATTTCAAC 529
Qy 246 IleValLeuThyArgBngLyGlyThrTyTyLeuArgPheLyPProThrAlaGlnGly 265
   |||||
Db 528 CTGCTGTCGCCCAACGCGCGCTCACTATATGACGCGCGCAAGCTGCGCGCGCTC 469
Qy 266 PheAlaGluThrLeuArgAenLeuSerGluIleSerProThrAlaTyLeuThrValPro 285
   |||||
Db 468 TTTCGACAGCTGCTGCGCAACCTGAAAGCGTGATGCGAGCTATTCAACGTGCGC 409
Qy 286 LyGlyTyTrpGluLeuValGlyAlaLeuGluArgPheThrLeuArgGluArgPhe 305
   |||||
Db 408 CGCGGCTTGACATGCTGATCGCGCATTCGCGCGTACCAAGCGTTGCGCGCGCTTC 349
Qy 306 PheAlaArgMetLyLeuPhePhePheAlaAlaGlyLeuSerGlnGlyIleTrpAsp 325
   |||||
Db 348 TTTCAGCAGGTGAAGTTCGCTTCTACGCGCGCGCACGCGCGCAACCTGTGGAC 289
Qy 326 ArgLeuArgPheValAlaGluGlnHisCysGlyGluArgIleArgMetAlaGlyLeu 345
   |||||
Db 288 GCGCTCGAACAATGTCGTCGAAACCGTCGCGCGCGCGCGCATGATCGGCGCTCG 229
Qy 346 GlyMetThrGluThrAlaProSerCysThrPheThrThrGlyProLeuSerMetAlaGly 365
   |||||
Db 228 GGTTCGACGGAACCTGCGCGTGGCGGACGACCTGATTCCTGCGCGAGCGCTGGC 169
Qy 366 TyIleGlyLeuProAlaProGlyCysGluValLyLeuValProValArgLyLyLeu 385
   |||||
Db 168 AATATCGCGCGCGCAACCGCGTACCGAGCTGAAGCTGTCACTCGCGCGCAACGCTG 149
Qy 386 GluGlyArgPheHisGlyProHisValMetSerGlyTyTrpArgAlaProGluGlnAsn 405
   |||||
Db 108 GAAGTGGCGCGCGCGCGCGCAACGTCACCGCGGTATCGGAAGCGCGCGAGCTGACC 49
Qy 406 AlAGlAlaPheArgPheGluGlyTyTrpCysSerGlyAspAlaIle 421
   |||||
Db 48 AGCGAGCGGTTTGACGACGAGGCGTTTATCTATCGGGAATGGCGTG 1

```

RESULT 2
 A2933866/c 671 bp DNA linear GSS 24-APR-2001
 LOCUS B2_Ba0001M12r B. japonicum BAC library Bradyrhizobium japonicum
 DEFINITION genomic, genomic survey sequence.
 ACCESSION A2933866
 VERSION A2933866.1 GI:13775926
 KEYWORDS GSS.

SOURCE
 ORGANISM
 Bradyrhizobium japonicum
 Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 Bradyrhizobiaceae; Bradyrhizobium.
 REFERENCE
 AUTHORS
 Tomkins, J.P., Wood, T.C., Stacey, M.G., Job, J.T., Judd, A.,
 Golcochea, J.L., Stacey, G., Sadowsky, M.J. and Wang, R.A.
 A marker-dense, sequence-ready map of the Bradyrhizobium japonicum
 genome
 TITLE
 Genome Res. 11 (8), 1434-1440 (2001)
 JOURNAL
 MEDLINE
 PUBMED
 COMMENT
 Contact: Ming RA
 Clemson University Genomics Institute
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: twing@clemson.edu
 Class: BAC ends
 High quality sequence start: 8
 High quality sequence stop: 625.
 Location/Qualifiers

FEATURES
 source
 1. 671
 /organism="Bradyrhizobium japonicum"
 /mol_type="genomic DNA"
 /strain="USD110"
 /db_xref="taxon:375"
 /lab_host="E. coli"
 /clone_lib="B. japonicum BAC library"
 /note="Vector: pIndigo536; Site_1: HindIII"

ORIGIN

Alignment Scores:

Pred. No.:	1,86e-34	Length:	671
Score:	435.00	Matches:	94
Percent Similarity:	57.92%	Conservative:	34
Best Local Similarity:	42.53%	Mismatches:	90
Query Match:	14.11%	Indels:	3
DB:	8	Gaps:	1

US-09-750-986d-30 (1-589) x A2933866 (1-671)

```

Qy 203 ThrGlaArgMetLeuCysAlaBngInMetLeuGlnThrPheProValPheGly 222
   |||||
Db 663 ACCACGCGAGTGTGACCTGACCGCACAGGCGCAAGCGCGACGACCTTCTCGAA 604
Qy 223 Glu-----GluProProValLeuValLeuPheProThrPhePheGly 240
   |||||
Db 603 CAGCGCGACGAGATCTGATCTTCGATCTGCTGCGCGCGCGACCACTTCCGCGCC 544
Qy 241 SerHisAsnIleGlyIleValLeuThyArgBngLyGlyThrTyTyLeuArgPheGly 260
   |||||
Db 543 AACCAACAATTTCAACTGCTGCGCGCAACGCGGCTCACTTATTCAGACGCGCGCAAG 484
Qy 261 ProThrAlaGlnGlyPheAlaGluThrLeuArgAsnLeuSerGluIleSerProThrAla 280
   |||||
Db 483 CTGCGCGCGCGTCTCTNCTGAGCTGCTCGCAACTGAAGCGTGATGACGAGCTC 424
Qy 281 TyLeuThrValProLyGlyTyTrpGluGluValGlyAlaLeuGluThrArgPheSerThr 300
   |||||
Db 423 TATTTCAGATGCGCGCGCGCTTTCGACATGCTGATCAGCGCATTTGCGGTGACGAAGCG 364
Qy 301 LeuArgGluArgPhePheAlaArgMetLyLeuPhePheAlaAlaIleGlyLeuSer 320
   |||||
Db 363 TTGCGCGCTCGCTTCTTCAGCGAGGAGTTCGCTTACCGCGCGCGCGCGCGCTGCGG 304
Qy 321 GlnGlyIleTrpArgPheArgLeuArgPheValAlaGluGlnHisCysGlyGluArgIleArg 340
   |||||
Db 303 CAGAACATGTGGCAGCGCGCTCAACTATGTCTCGTGTAACTGCGCGCGCTGTGCGCG 244
Qy 341 MetMetAlaGlyLeuGlyMetThrGluThrAlaProSerCysThrPhe-ThrThrGlyPhe 360
   |||||

```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 7, 2005, 23:57:11 / Search time 4035 seconds
(without alignments)
5319.207 Million cell updates/sec

Title: US-09-750-986D-30
Perfect score: 3082
Sequence: 1 MRSLEALLPFGRIERLEH.....KVDALYRGEDQSMIRDEATL 589

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 32822875 segs, 18219865908 residues
Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xid
-O/cgpt 1/USPTO.spool.p/us09750986/runat.03022005.071744.18577/app.query.fasta.1.775
-UNITS=bits -START=1 -END=1 -MATRIX=bl0sum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pcio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09750986 @CGN 1.1 5180 @runat.03022005.071744.18577 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEJOURNEY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGIO
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: EST:
1: gb_est1:*
2: gb_est2:*
3: gb_hic:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	514	16.7	708	8	AZ934119	BJ_Ba000
C 2	435	14.1	671	8	AZ933866	BJ_Ba000
C 3	315	10.2	714	8	AZ933744	BJ_Ba000
4	301.5	9.8	544	8	AZ935488	BJ_Ba000
5	271	8.8	2698	3	AK054103	Mus muscu
6	264.5	8.6	2505	3	AV109410	Zea mays
7	255	8.3	2848	3	AK004897	Mus muscu
8	253	8.2	1897	9	AY402446	Mus muscu
9	253	8.2	2548	3	CR606980	full-leng

10	250	8.1	2518	3	BC026161	BC026161 Homo sapi
11	244	7.9	2097	9	AY402444	AY402444 Homo sapi
12	244	7.9	2203	3	CR597120	CR597120 full-leng
13	232.5	7.5	2672	3	AK028078	AK028078 Mus muscu
14	226	7.3	2175	9	AY418305	AY418305 Homo sapi
15	226	7.3	2366	3	CNS0A7CK	EX824843 Arabidops
16	220	7.1	4243	3	AK034942	AK034942 Mus muscu
17	215	7.0	2166	9	AY418307	AY418307 Mus muscu
18	211.5	6.9	979	7	CK248477	CK248477 EST732114
19	201.5	6.5	1199	6	CD508894	CD508894 CDA93-D04
20	198.5	6.4	907	9	CG847891	CG847891 ZMMBBD031
21	197.5	6.4	2461	3	BC028399	BC028399 Homo sapi
22	197	6.4	829	6	CB675484	CB675484 OSUNE11E
23	197	6.4	831	6	CB675482	CB675482 OSUNE11E
24	194.5	6.3	909	7	CO114394	CO114394 GR_EB015
25	194.5	6.3	941	5	BX705748	BX705748 EST706492
26	190.5	6.2	898	7	CK260414	CK260414 EST706492
27	187.5	6.1	815	7	CO127557	CO127557 GR_EB12D
28	186.5	6.1	1034	8	B2676934	B2676934 PUBIF45TD
29	186.5	6.1	1265	6	CD503756	CD503756 CDA64-B06
30	185.5	6.1	2290	3	CNS0AARU	BN157496 946208 MA
31	185.5	6.0	909	7	CN157496	CN157496 946208 MA
32	184.5	6.0	766	6	CD445611	CD445611 EL01T0404
33	183	5.9	1811	3	CNS0AD10	EX818812 Arabidops
34	183	5.9	2977	3	AK012088	AK012088 Mus muscu
35	182	5.9	1833	3	CNS0A10J	EX832101 Arabidops
36	181	5.9	884	7	CK260415	CK260415 EST706493
37	180.5	5.9	810	7	CO200299	CO200299 GE02.6.D0
38	180.5	5.9	1841	3	CNS0ADDC	EX813547 Arabidops
39	180	5.8	1168	5	BX344418	EX344418 BX344418
40	179	5.8	931	7	CK245466	CK245466 EST729103
41	177.5	5.8	901	7	CK800668	CK800668 AGENCOURT
42	176.5	5.7	1920	3	CNS0A5CA	BX823446 Arabidops
43	176	5.7	796	4	BJ719420	BJ719420 BJ719420
44	175.5	5.7	1200	6	CD506373	CD506373 CDA78-H04
45	175	5.7	1832	3	CNS0AD12	EX814596 Arabidops

ALIGNMENTS

RESULT 1
AZ934119/c
LOCUS AZ934119 708 bp DNA linear GSS 24-APR-2001
DEFINITION BJ_Ba0001H18f B. japonicum BAC library Bradyrhizobium japonicum
ACCESSION AZ934119
VERSION AZ934119.1 GI:13776179
KEYWORDS GSS.
SOURCE Bradyrhizobium japonicum
ORGANISM Bradyrhizobium japonicum
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Bradyrhizobiaceae; Bradyrhizobium.

REFERENCE 1 (bases 1 to 708)
Tomkins,J.P., Wood,T.C., Stacey,M.G., Loh,J.T., Judd,A.,
Goicoechea,J.L., Stacey,G., Sadowsky,M.J. and Wing,R.A.

A marker-dense, sequence-ready map of the Bradyrhizobium japonicum
Genome
Genome Res. 11 (8), 1434-1440 (2001)

TITLE JOURNAL
MEDLINE 21376150
PUBMED 11483585

COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: twing@clemson.edu

Class: BAC ends
High quality sequence stop: 661.
Location/Qualifiers

FEATURES
source 1..708
/organism="Bradyrhizobium japonicum"
/mol_type="genomic DNA"

Db 361 ECEP--LVYGNPFDLTITGTAGIGPVSTREIVYDDAGNVVNDQVVELVGRGQWQGY 418
QY 399 WRAPENQAQAFDEBGYYCSGDAIKLAD-----PADPOKGLMFD 436
Db 419 WQPEATKEVYLNAGMWTSDIVKFDQGLIHIVRKKDMILVSGFNVYFNEIEDVALH 478
QY 437 GRAEDPKLSSGVVSGPIATRAVLEGSGVLDVVVAAPDRBCLGLVFPRLDGR--- 493
Db 479 GKVLVNAIQANDVSGELVYKIVYKRPDLTKDEVIAGCKRKLITGYKV--PKLVFRRDL 537
QY 494 ALSGLGK-----EASDAEVL 509
Db 538 PKTNVGIKLRVLRRENDQDLAA 560

RESULT 15

E70937
probable fadD15 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Jul-2004
C/Accession: E70937
R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998
A/Authors: Sgaree, R.; Suleston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A/Reference number: A70500; MUID:98295987; PMID:9634230
A/Accession: E70937
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1600 <COL>
A/Cross-references: UNIPROT:O53521; GB:AL021957; GB:AL123456; NID:G3242293; PID:CAA1749
A/Experimental source: strain H37RV
C/Genetics:
A/Gene: fadD15
C/Superfamily: Synchocytis long-chain-fatty-acid-CoA ligase; acetate-CoA ligase homoid
F/66-585/domain: acetate-CoA ligase homology <ACL>

Query Match 8.4%; Score 257.5; DB 2; Length 600;
Best Local Similarity 21.7%; Pred. No. 3.3e-12;
Matches 132; Conservative 101; Mismatches 267; Indels 109; Gaps 22;

QY 34 RAANGERRISYEMFNVAIAQSLLPYGLSARPLIYSGNDLEHQLAFGAMVAGIP 93
Db 37 RLIDGVWTDVTCAPAAQIRPAAALGLISGVQAGDRVIFSAITREYEMALIDFALAVG-- 94
QY 94 YCPVSPAYSLSDOLAKLRHIVGLQGLVFAADAPFORAIEITILPDVPAIFTRGELA 153
Db 95 -AVTVPTERTSSAE--QVRVLODSEAVVLPFAETDSHATWAE--LSGSVPALREVLOIA 149
QY 154 GRRVTSFDSLLEOPGIEADN-----APATGPPTIAKFLFTSGSTKLPKAVPTQRL 207
Db 150 GSGGNALDRLTEAGASVDPALFTARLAALRSTDPATL---IYTSGTGRPKGCQLTQSNL 206
QY 208 CANQOMLQTFPVGCEPPLVDMPLPMNHTFGSGSHNIGIVLYNGGTYYLDDGKPTAQGFA 267
Db 207 VHEIKGARAVHPTLLRGERLVLPLAHVL--ARAISMAAFH-----SKYTV--GFT 255
QY 268 ETLRN-----LSEISPTAYLVTPKGMELVGLERDSTLRERFFARMKLPFAA----- 316
Db 256 SDIKNLLPMLAVKFTVVSVPKVFKEKYNVTAEGNAANA---GKGRIFALMAQTAVDWS 311
QY 317 -----AGL-----SGIWMRL-----DRVABQHGERIRM 341
Db 312 EACDRGCGPGLLRAKHAVPDLVTRKLRALGNCRAAVSGAPLGARLGIFYRGAGLTI 371
QY 342 MAGIGMETAPSCFTTGPLSMAGYIGLPAPGCEVXLVPVDGKLEGRFHHVSGYRA 401
Db 372 YEGYISGTSIGVAISQFNDIKIGTVGKVPFNGSLRIAD--DGEILVR--GQVVFSGYRN 428
QY 402 PEONAAQAFDEBGYYCSGDAIKLADPADPOKGLMFDGRIABFKLSSGVFVSGPLRTAV 461

Db 429 EQATTEAF--TDGWFXTGD-----LGAVDEBGFLLITTRKKEIIYTAGG--KNVAPVLEDO 481
QY 462 LEEGSYVLDDVVAAPDRBCLGLVFPRLDCLASG-----LGEASDAEVL 509
Db 482 LRAHPILSQAVVVGDAKPFIGALI---TIDPEAFEGMKGRNSKTAGASVGDLDATDPLA 538
QY 510 SEVVRAMFADWLKRLREATGNSRIIMWVGLDTPPSIDKGEVTDKGSINORAVLQWRS 569
Db 539 EIDAA-----VKQANL-AVSHAESIRKPRILFVDFTEDTGELTPMKVARKVAEKFA 591
QY 570 KYDALYRGE 578
Db 592 DIEAIYNKE 600

Search completed: February 7, 2005, 23:50:01
Job time: 42 secs

Db 148 SIRVOLLANDSGEHNLODAINHPADGTAAPPSPADDEVVYFQLSGTTGTPKLIPTRTNDY 207
QY 205 -----RMLCANQOMLLQTFPVGGEPPVLDVLPNNHTFGSGH--NIG 245
Db 208 YGVRVRSVEICQFTQCRYLCA-----IPXHNHYANSSPGSLG 245
QY 246 IVLNGSTYVLDGKPTAOGFAETL-----RNLSISPTAY-----LTVPKMEE 290
Db 246 VFL-AGGTAVL--AADPSA-----TLCFPLIEKHQINVTALVPVAVSLMLQALTEGESRAQ 298
QY 291 L-----VGALEBRDSTLRERFFARMKLFPPAAAGLSOGI--WDRLDVRVAEOHGERIRM 341
Db 299 LASLKLQVGGARLSATLAARIPAEICQLOQVFGMAEGLVNTYRLDLSAEK----- 350
QY 342 MAGLGMTETAPSCFTTTPPLSMAGYIGLPAPGCEVVKLVDPDGKLEGFHGFHWMSGYRA 401
Db 351 ---IHTQGYPMC--PDDEVWVADAEENPLPQGEV-----GRLMTR--GPYTFRGYKS 397
QY 402 PEONAOAFDEEGYCGSDAIKLADPADPQKGLMFDGRIADDFKLSSGVFVSVGLRTTRAV 461
Db 398 POHNASAFDANGFYCSGDLISI-----DREGYITVQGREKQ-----I 435
QY 462 LEGGSYLDVVVAPADRECGLVFPRLDORALS---GLKESADAEVLASEPVRAMFA 518
Db 436 NRQGE-----KIAAEETENL-LRHHPAVITAAVLSMEDELMEGKSCAYLVVKEPLR---A 486
QY 519 DMLKRLNREATGNASRIMVGLDTPPSIDKEVTDGSGINORAVLQWRSAKYDA 573
Db 487 VQVRRLFRE-----QGIAEFKLPDRVCEVDSLPLTAVGVKDKKQRLQWLASRASA 536

RESULT 13
A99708
2,3-dihydroxybenzoate-AMP ligase [imported] - Escherichia coli (strain O157:H7, substra
C:Species: Escherichia coli
C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C:Accession: A99708
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
gaeswara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: A99708
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-536 <HAV>
A:Cross-references: UNIPROT:Q8XBVJ; GB:BA000007; PIDD:BA034056.1; PID:913360091; GSFD8:G
A:Experimental source: strain O157:H7, substrain R1MD 0509952
C:Genetics:
A:Gene: ECG0633
C:Superfamily: 4-coumarate-CoA ligase; acetate-CoA ligase homology

Query Match 8.4%; Score 260; DB 2; Length 536;
Best Local Similarity 24.0%; Pred. No. 1.8e-12;
Matches 143; Conservative 77; Mismatches 215; Indels 160; Gaps 27;
QY 37 NGEWRISYAEHFNVAIAQSLIPYLSAERPLLVSGNDLEHQLAFGAMTAGIYCP 96
Db 44 DGB-RQSLYRELNOADNLACSLRQGIKRGFTALVGLNAVLEYITFFALKLG----- 97
QY 97 VSPASVLLSODLAKLRHIVGLQGLVFAADAAPFORAIEITLPPDVPAIFRTGELAGR 156
Db 98 VAVVLLALFHSQRSELNVAASQIEPALIIA-----DRQHALLFGSD-----FLMTFVAEHS 147
QY 157 TVSFDSLLEPG-----GIEADNAFAATGP-DTIAKLVFTSGSKLPRKAVPTTQ--- 204
Db 148 SIRVOLLANDSGEHNLODAINHPADGTAAPPSPADDEVVYFQLSGTTGTPKLIPTRTNDY 207
QY 205 -----RMLCANQOMLLQTFPVGGEPPVLDVLPNNHTFGSGH--NIG 245
Db 208 YGVRVRSVEICQFTQCRYLCA-----IPXHNHYANSSPGSLG 245
QY 246 IVLNGSTYVLDGKPTAOGFAETL-----RNLSISPTAY-----LTVPKMEE 290

Db 246 VFL-AGGTAVL--AADPSA-----TLCFPLIEKHQINVTALVPVAVSLMLQALTEGESRAQ 298
QY 291 L-----VGALEBRDSTLRERFFARMKLFPPAAAGLSOGI--WDRLDVRVAEOHGERIRM 341
Db 299 LASLKLQVGGARLSATLAARIPAEICQLOQVFGMAEGLVNTYRLDLSAEK----- 350
QY 342 MAGLGMTETAPSCFTTTPPLSMAGYIGLPAPGCEVVKLVDPDGKLEGFHGFHWMSGYRA 401
Db 351 ---IHTQGYPMC--PDDEVWVADAEENPLPQGEV-----GRLMTR--GPYTFRGYKS 397
QY 402 PEONAOAFDEEGYCGSDAIKLADPADPQKGLMFDGRIADDFKLSSGVFVSVGLRTTRAV 461
Db 398 POHNASAFDANGFYCSGDLISI-----DREGYITVQGREKQ-----I 435
QY 462 LEGGSYLDVVVAPADRECGLVFPRLDORALS---GLKESADAEVLASEPVRAMFA 518
Db 436 NRQGE-----KIAAEETENL-LRHHPAVITAAVLSMEDELMEGKSCAYLVVKEPLR---A 486
QY 519 DMLKRLNREATGNASRIMVGLDTPPSIDKEVTDGSGINORAVLQWRSAKYDA 573
Db 487 VQVRRLFRE-----QGIAEFKLPDRVCEVDSLPLTAVGVKDKKQRLQWLASRASA 536

RESULT 14
H82132
long-chain-fatty-acid-CoA ligase VC1985 [imported] - Vibrio cholerae (strain N16961 serog
C:Species: Vibrio cholerae
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
C:Accession: H82132
R:Heidelberger, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gynn, M.L.; Dodson, R.J.;
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Baas, S.; Qin, H.; Dragol, I.; Sellers, P
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Frazer, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: H82132
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-566 <HEI>
A:Cross-references: UNIPROT:Q9KQJ3; GB:AE004273; GB:AE008527; NID:9656517; PIDD:AAE9513;
A:Experimental source: serogroup O1, strain N16961; Biotype B1 Tor
C:Genetics:
A:Gene: VC1985
A:Map position: 1
C:Superfamily: 4-coumarate-CoA ligase; acetate-CoA ligase homology
F:70-550/Domain: acetate-CoA ligase homology <ACL>

Query Match 8.4%; Score 258; DB 2; Length 566;
Best Local Similarity 24.3%; Pred. No. 2.7e-12;
Matches 122; Conservative 69; Mismatches 218; Indels 94; Gaps 18;
QY 76 NDEHQLQAFGANVAGIIPYCPVSPAYS-----LISODLAKLRHIVGL 118
Db 83 NLLQYVALFGVIRACMIANVAPLTPPRELEHQLNDADRAIVIVSNFANTLEQIVANT 142
QY 119 Q-EGVFAADAAPFORAIEITLPPDVPAIFTRG-----ELAGRTVSFDLSLEPGGIEA 172
Db 143 QVGHVVLTSIGWLPRAKGIIV--DVVVKYKGMVRYKIDLPG--AISMRAALKHGRRLQY 198
QY 173 DNAFAATGPDITAKLFTSGSTKLPRVPTTQRMLCANQOMLLQOT---PVGGEPPVL 228
Db 199 VKPF-MSGED-IAPLQYTGTTGVAAGAILTHRNVAAN---VLQAKGAYFPVLOEGRELV 253
QY 229 VDMLPNNHTFGSGSHNIGIVLVNGITYVLDGKPTAOGFAETLRNLSEISTAYLVTPKGM 288
Db 254 VTLAPYHVFALVNLCLFLEMGRMLITNPBDIGFVKELOKYP-----FTAITGV 306
QY 289 EELVGALEBRDSTLRERFFARMKLFPPAAAGLSOGIDRLDRVAEOHGERIRMAAGLGT 348
Db 307 NTLFNLVNNEDHELDPRMKLAVGGMVQGAIVAEKMKTTGVH-----LLEGYGLT 360
QY 349 ETAPSCFTTTPPLSMAGY---IGLPAPGCEVVKL-----VPVDGKLEGFHGFHWMSGY 398

Db 202 -----ANVIVDT---OKOLEKILKIMWDLPHLKAUVIYQEPPEKMANVYTMELI 250
QY 154 GRRTVFDLSLEPGGIEADNAPAAAT-GPRTIAKFLFTSGSTKLPKAVPTTQRMCLC----- 208
Db 251 E-----LGEVPEBEALDAILITQOPNOCVLYVTSGTTGPKVMLSODNIWTAR 301
QY 209 -ANQOMLLQTFPVGEEPPVLDMLPMNHTFGSGSHNIGIYLYNGGTYYLDDGKPTAAGFA 267
Db 302 YGQADGIQPAEV---QGEVVVSYLPLSHIAAQIYDLMGTGQKAOUCFPADPAKCTIV 358
QY 268 ETLRNLSISPTALVLPKGMELVGLERDSTLERFPFAMKLPFAAGLSGI----- 323
Db 359 NTLR---EVEPTSHMGVPRVWEKIMERIQ-EVAAGSGFIRKMLMAMSVTLQNLTPS 414
QY 324 -----MDRLDRVAEQ-----HCGEIRIMMAGIGMT 348
Db 415 NDKPFTSLADYLVLARVROALGPAKQKXNFGAAMTAEOTRFLGLNIRLPAAGLS 474
QY 349 E-TAPSCFTTGPLSMAGY-IGLPAPGCEVVLV--PVDGKLEGRFHGPHVMGTYRAPEQ 404
Db 475 ESTGPH--FMSSPYNVLYSSGRVPCRYKLVNQDDAGICEICLMKRTTFMGILNMDK 532
QY 405 NAOAFDEEGYCCGDAIKLADPADPQKGLMFDGRIADFKLSGVPVSVGLRTRAVLEG 464
Db 533 TCERIDSEGMVHTGDMQRL-----DADGFLYITGRKELIITAGGENVPVPIEBAVME- 587
QY 465 GSIVLDVVAA-----PRECLG--LVFPRLLD-----GRALSGLOKEA 502
Db 588 ----LPIISSAMLIGDQKFLSMILTKCTLDPTSEPTDSLREQAVEFCORVSKASTV 643
QY 503 SDAEVLASEPVAAMFADMLKRLNREATGNASRIIMVGLDTPPSIDKGEVTDKSGINQRA 562
Db 644 SEIVGQDEADVQAIHIGICRVANNAAPRYHIQKMAIILQRPDSISGSELGPTMKRLT 703
QY 563 VLQMSAKVDALYRGEDQ 580
Db 704 VLRRYKDIIIDSFYQEQKQ 721

RESULT 11

H69274
Probable long-chain-fatty-acid-CoA ligase (EC 6.2.1.3) AF0200 [similarity] - Archaeoglob
C/Species: Archaeoglobus fulgidus
C/Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004
C/Accession: H69274
R/Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kittness, E.F.;
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A/Authors: Overbeek, T.; Cotton, M.D.; Spriggs, T.; Artlisch, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A/Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae
A/Reference number: A69250; MUID:98049343; PMID:9389475
A/Accession: H69274
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-601 <KLE>
A/Cross-references: UNIPROT:O30039; GB:AE001092; GB:AE000782; NID:G2689415; PIDD:AA9103
C/Superfamily: Synecocystis long-chain-fatty-acid-CoA ligase; acetate-CoA ligase homolo
C/Keywords: acid:thiol ligase; coenzyme A
F:65-587/Domain: acetate-CoA ligase homology <ACL>

Query Match 8.6%; Score 265; DB 2; Length 601;
Best Local Similarity 21.4%; Pred. No. 8.4e-11;
Matches 131; Conservative 106; Mismatches 253; Indels 122; Gaps 21;

QY 37 NGEWRRISSVEMFNHRAIAQSILPYGLSARPLLIVSGNDLEHLQALFAMVAGIPYCP 96
Db 39 DGRKLVTYEFWRVWRKSLFSSGLRGKGRDAIVADRYWEIADPAVLTAG---GV 95
QY 97 VSPAYSLSDLAKLRIHIVGLQPGVLPADAAFPQAIETILPDDVPAIFTRGELAGR 156
Db 96 VVTVHSLVNRB--QVEYILRDSRSRVVFT-----EKYAEVNDPEDEVFLB----- 140

QY 157 TVSFDSLLEPGGIEADNAPAAATGPRTIAKFLFTSGSTKLPKAVPTTQRMCLC----- 216
Db 141 --ELEKLVGEVSDDEESRKKSVEPDDLASIVTTSGETGPKAMLTMMWRNCHSVMS 198
QY 217 TFPVGEPPVLDMLPMNHTFGSGSHNIGIYLYNGGTYYLDDGKPTAAGFAETLRNLSEI 276
Db 199 ITTFYGEPHIC--YLPLSHVQ-----RLVFPAGISRAATAVFCSPQOPLFT---STAV 248
QY 277 SPTAYLTPKGMEE---LVGALERDSTLERFPFAMKLPFAA-----AAGLSQ 321
Db 249 KPVGLVPRILRIVNAGIVEKEKSPALAK-----KIFWARGVAIEGKMSRGEKY 302
QY 322 GIV-----DRL-----DRVAEGCHGERIRMMAGIGMT 348
Db 303 GFWLNKRIIADLVSKIRENLGLTRIRFVCSAAELQELAYMNGKIPVIEGIGMT 362
QY 349 ETAPSCFTTGPLSMAGYIGLPAPGCEVVLV--PVDGKLEGRFHGPHVMGTYRAPEQ 408
Db 363 ETAPSNLNVGRFPKQTVGPPIRGIE-EAIAEDGEILVR--GDNVKMGYMRKEATRTKT 419
QY 409 FDEEGYCCGDAIKLADPADPQKGLMFDGRIADFKLSGVPVSVGLRTRAVLEGSYV 468
Db 420 FTEDGMLKTGDLGEF---DEDGVLVFLGRKKHIIYLDTGKNVSPVIEEE--LKNPLV 473
QY 469 LDVVVAAPDECGTGLVFPF--LIDCRALSGL-----GKE---ASDAEVL 508
Db 474 SDAVITIGDGKPYTTILVFPFSLFADKNGIEYRSRTIVORSISGEIYAVDENLV 533
QY 509 ASEPVRAWA---DWLKRINREATGNASRIIMVGLDTPPSIDKGEVTDKSGINQRAVL 564
Db 534 ENNAVVELYAKIVDVNSRLAKHET-----IKFKILPEAFSLKGEIPTLKKRRHVL 588
QY 565 QMSAKVDALYR 576
Db 589 KRYEKPIEMVYK 600

RESULT 12

E85558
2,3-dihydroxybenzoate-AMP ligase (imported) - Escherichia coli (strain O157:H7, substrain
C/Species: Escherichia coli
C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 01-Mar-2002
C/Accession: E85558
R/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,
Hiller, L.; Grobbeck, B.O.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A/Reference number: A85480; MUID:21074935; PMID:11206551
A/Accession: E85558
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-536 <STO>
A/Cross-references: GB:AE005174; NID:G12513487; PIDD:AA654929.1; GSPDB:GN00145; UWGP:207;
A/Experimental source: strain O157:H7, substrain EDL933
C/Genetics:
A:Gene: entE
C/Superfamily: 4-coumarate-CoA ligase; acetate-CoA ligase homology

Query Match 8.5%; Score 261; DB 2; Length 536;
Best Local Similarity 24.0%; Pred. No. 1.5e-12;
Matches 143; Conservative 77; Mismatches 215; Indels 160; Gaps 27;

QY 37 NGEWRRISSVEMFNHRAIAQSILPYGLSARPLLIVSGNDLEHLQALFAMVAGIPYCP 96
Db 44 DGE-RQISTRELNQAADNLCSLRQGIKRGETALVQGVAVLYTTPFALLKG----- 97
QY 97 VSPAYSLSDLAKLRIHIVGLQPGVLPADAAFPQAIETILPDDVPAIFTRGELAGR 156
Db 98 VAVVLAHLSHQRSBLNVAVSAQIEPALLIA-----DRQNALPFGDD-----FLMTFAEHS 147
QY 157 TVSFDSLLEPGG-----GLEADNAPAAATGP-RTIAKFLFTSGSTKLPKAVPTTQ--- 204

F556-583/Domain: acetate-CoA ligase homology <ACL>

Query Match	9.2%;	Score 283.5;	DB 2;	Length 601;
Best Local Similarity	21.8%;	Pred. No. 2.9e-14;		
Matches 136;	Conservative 94;	Mismatches 252;	Indels 143;	Gaps 24

```

Qy      37 NGEMRRISYAMFHNVRALIAOSLI.PYGASARPLLIVSGNDLEHQLAFCGMVAGIPYCP    96
Dz      30 NGOMOAILWHOPGEOLDELISWALLAOIGIVODKTAIFSNMPPRTIADFAALOIRATVP    89
Qy      97 VSP-----AYSLSODLAKRHHVGLLOGLVFPAADAFAFORAIETILPDDVPAFTFR    149
Dz      90 IYPTNTPBQAAYIIIONADVK-----VVFGBAQAFDALISGF--EQCBEL--R    133
Qy      150 GELGARRVVS-----FDLSLEOPGGIEADNAFAATGPDTIA    185
Dz      134 LIVMANANIDIKOASCAMHWDEFVAOSHODRAPLDALIEQ-----ANYDLFL    181
Qy      186 KFLFSTGSYTKLPKAVPTTORMICANOQMLOTTFPFGESEPVULVDWLPMNHTPGSHNIG    245
Dz      182 TLITTSCTTGTPKGVMLDYRNIGAOBGHORLINTQED--VSLCFPLSHVERAWT--A    238
Qy      246 IVLYNGGT--YYLDDCKPTAOGFAETLRNLSEISPTAYLTVEKGBEVLGALERDSTURER    304
Dz      239 YVLKYKGAICYLOD-----VAHVRELAIEVRFTVMCAVPRFEKIPISAIH--EKVKA    289
Qy      305 PFAPMKLFPPAA-AGLSGI-----W-----DRL-----DRVAEHC    335
Dz      290 PLVKRVKLVTAMVNNGAKMAVCROOORPSMWLKOSHOFADKLVIUSKLRALLGRINFWPC    349
Qy      336 GER-----IRMMAGLMETTASCTFTTGPLSMAGYIGLPARGCEVKLVPV    381
Dz      350 GGALUDETIGFFHAIGINVLGYGMTETTATVSCMDHCHNPISIGISMGAQVK---I    406
Qy      382 DGKLEGRFHGPHVNSGWIRAPEONAQAFDEEGYVCSGDAILADPADPOKLMFDRIAE    441
Dz      407 GENNEILLVRGEMVWRGYXKDKETTESFEOGLFKTAGCHI----DENGMFITDRIKE    462
Qy      442 DFKLSSGVFVSVGLRTRAVLEGG---SYULDVVNVAAPDEECGLVFPRLDLC--RALS    496
Dz      463 LMKTSGGXYIA-----PQVTEGAIGKDHFLEQIAVLINDTKKFVSALIVP---CPSDLE    512
Qy      497 GLGKEAS-----DAVLASEBPVRAMFADWLRLRLREATGNARKIMWVGILTDPESIDKE    551
Dz      513 EYAELINKYHDRLIELIKHSQVLTEMFEKRVMELQKE--LAKEQYVKRPFLPLPKASMSGE    571
Qy      552 VTDKGSINORAUVLOWRSAKVDALYR    576
Dz      572 LTPTOKRRRKVINDRYODEIEWMO    596

```

RESULT 9

T07944

probable long-chain-fatty-acid-CoA ligase (EC 6.2.1.3) - rape
C. Species: *Brassica napus* (rare)

С. Ю. БРАДОВА (Ташкент)

```
CjDate: 21-May-1999 #sequence_revision 21-May-1999 #text_change 09-Jul-2004
```

C;Accession: T07944

R; Fulda, M.

submitted to the EMBL Data Library, April 1996

A;Reference number: Z16223

A;Accession: T07944

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-677 <FUL>

A: Cross-references: UNIPROT:096338; EMBL:Z72152; PIDN:CAA96522.1

A: Experimental source: cv. Ascaris: developmental stage: 20-26 days after flowering: tissue

C:Superfamily: human long-chain-fatty-acid-CoA ligase; acetate-CoA ligase homolog

Keywords: acid-thiol lyase; coenzyme A

F:135-670/Domain: acetate-CoA ligase homology <ACL>

Query Match 8.9%; Score 273; DB 2; Length 677;
Best Local Similarity 24.5%; Pred. No. 2.3e-13;
Matches 121; Conservative 74; Mismatches 175; Indels 124; Gaps 24;

```

QY 38 GEMRISVAEMHNRAALAAQSLPLPGLSMERPLLIYSGDLEHL---OLAFGAMVYGIPY 94
Db 110 GDYKMMTIGEVTATLALGSLGVHHGITTGSSVGCIYFIRPEMLIYDQACASISYSV-- 167
QY 95 CPVSPAYSLLSODLAK--LRH-----IVGLLQGLVFAADAAPFORAIETI--LPDDV 143
Db 168 ----FLUYPTLGDADAKFYVNHANVOAIFCVAETLSLISGLSEMPSEVRLLVVVVGINESL 223
QY 144 PAIFIRGELAGRRTVSPDSLEPOGIEADNAFAATGPTIKFLFTSGSTKLPKAVPT 203
Db 224 PSL---PPSAGYKAVSYSLVNLQ--GRSNPOFPCCPKPDVATLCYTSGETGPKGVLT 278
QY 204 OMLCANOQMLLQTFEPVFGEEPVLVDWLPWNHTFGSGHNIGVLYNG---GTYLLDGG 260
Db 279 HANLIAN--VAGSSSEYKFFSSDIYISLPLHIVERANOI--LAVFGVAVGEYQGDNMK 335
QY 261 PTAQGAETLRVLSLSPAYLTVP---KGMELVGALEKSTURERPF---ARMKL 311
Db 336 ----LDDLLALRPVFSSEVPLINRYITDGTINAAVKTSGGLKERLPNAAVNAKKA 387
QY 312 PFFPAAGLSOGIWRDL-----DRVAEOHCGERIRMA----- 343
Db 368 LLNGKS--ASPIIDRLVFNKIKIDL-----GGRVRFMTSGASPLSPEVLEFLKICFGRY 440
QY 344 --GLAMETAPSCFTT---TGPLSMAGYGLYAPGCEVTVLPVD-----GKL 385
Db 441 SEGVMETEL--SCVISGMEDEG-NLTGHGSGPNPADEVLTMDVPEMNYTSADEPHRGEL 497
QY 386 EGRFEGPVMYSGYMRAPRONAQAPEDEGGYCCGDAIKLADPADPQKGLMFDG-----R 438
Db 498 CVR--GPILFPGYKDEOVQTRIVEDDGLHTGDI-----GLWLPGRLLKTIIDR 544
QY 439 IADPKLSSGVFVS 452
Db 545 KKNIFKLAQGEYIA 558

```

RESULT 10

JC7557

lipidosin - moube

N;Alternate names: Lipidosis-related protein

C;Species: Mus musculus (house mouse)

C;Date: 30-Jun-2001 #sequence revision 30-Jun-2001 #text change 09-Jul-2004

C;Accession: JC75557

R;Moriya-Sato, A.; Hida, A.; Inagawa-Ogashiwa, M.; Wada, M.R.; Sugiyama, K.; Shimizu, J.

Biochem. Biophys. Res. Commun. 279, 62-68, 2000

A:Title: Novel acyl-CoA synthetase in adrenoleukodystrophy target tissues.

A:Reference number: JC7557; MUID: 20563802; PMID:11112418

A: Contents: Brain

A: Accession: JG7557

A: Molecule type: MF

A: Residues: 1-721 <MOB-

A:Cross-references: INTBROT:089PITE
H/KEBUCEB: 1-721 (NOV)

C.Comment: This protein possesses long-chain acyl-CoA synthetase and catalyzes

cyclo-oxygenase activity and catalyzes the conversion of fatty acids to eicosanoids. This

tion and the impairment of fatty acid in-

in human X-linked adrenoleukodystrophy

Query Match	8.8%	Score 270;	DB 2;	Length 721;
Best Local Similarity	22.7%;	Pred. No. 4.4e-13;		
Matches 154; Conservative	87;	Mismatches 287;	Indels 150;	Gaps 24

```

0Y      12 GRILRLHMAKTRP---EQTCLVBARANG-----EMRRLSYAEMFNINPAIAO 57
      85 GRVRLRLPPTCTORYTYVHOMFYELDLKRYGNLSALGFKRKDMERISYQYLLIARVAK 144
0Y      58 SLLPYGLSAERPLLVSGNDLEHLTLAGAMVAG---IPYCPVSP-AYSLSQDLAKR 112
Db      145 GFLKGLGLRAHSAVALGNSPFWPSAAGTVAFAGGIYVIGYITTSSEACQYISHDR--- 201
0Y      113 HIVGLQGLVPAADADPFOAIEFTIL-----PDDVPAIFTRGELA 153

```


C:Genetics:

A:Gene: YP00537

C:Superfamily: Synecocystis long-chain-fatty-acid-CoA ligase; acetate-CoA ligase homolo

Query Match 9.4%; Score 289; DB 2; Length 601;

Best Local Similarity 21.3%; Pred. No. 1,1e-14;

Matches 135; Conservative 115; Mismatches 268; Indels 116; Gaps 22;

```
QY 14 ILERLHMAKTRPQTCVAAARAANGEMRRISYAEFNNVPAIAQSLIPYGLSARPLLIY 73
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 10 LVRLRLQOVSDRSELVAFRWSPGE-KQLNWQIDHIVTRISYALLSLAAIOERIGIF 68
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 74 SGR-----DLEHLQAFGAMVAGIPYCVSPAYSLSDLAQRHIVGLQGLVFAA 126
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 69 ANNSMAMSLVDLAILDL-----RGVSV-----PLIYA--TNTAAQAYIVDADVRILFVG 116
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 127 DAAPFORAI--ETILPDDV-----PAIFTRGELAGRTVSPSLERQPG----- 168
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 117 EGQOFDAITLQPLCQLAHIIIVLDPVADLRGCEYAGHLADFEG-QQOPDTVOQHLLTTR 175
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 169 --GLEADNAFAAGPDTIAKFLPFGSGTKLPKAVPTTORMLCANQOMLQTFPVGSEEP 226
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 176 IECDDLDLFT-----LIYSGTGTGEPKGMLDYRMAAQLYLHDRLTLTPD-- 224
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 227 VLVDMLPMNHTFGSGSHNIGIVLVNGGTYYLDDGKPTAQFAETLRN-LSEISPTAYLTV 285
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 225 VSLSFLPLSHVFERAMSPYVMTGAQNVYTRN-----TDVYRSAMQAVKPTVMCAVP 276
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 286 KGMEEVLGALERDSTLEREPFARMKLPFAAG-----LSQGIWDRLDYV 330
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 277 RPYEKVPSAI-NDKVALAKMHQRYLPHMAVGCGRKPKLQOGRPLPMLSEQMTTLADRL 335
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 331 AEQH-----CGERIRMA-----GLGTEPARPCTFTTGLSM 363
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 336 VLKRGVLGGRVRFLEPAAGALDDNIIILFQALGVNIKKYGYGTEFCATVSCWEEQNR 395
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 364 AGYIGLPAPGCEVVLVVDGKLEGRFHPHWSGYWRAPBONAOAPDEGYVSGDAIKL 423
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 396 FGSIGKRLPGIDRL--GAENEIQAQVGYVMRGYFKKPOETLEAFEDGMLKTGDGAL 452
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 424 ADPAPOKGLMFDGRIADDFKLSGVFVSVGPRLTRAIVLGGSYVLDVVVAADRECLGL 483
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 453 ----DAQHLLFITERLKDLMKTSQKRYIA--PQWIEGTLGDRPIEQVALIADTRKVSVA 506
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 484 LVPRRL--LDCRALSGIKASDAEVLASEPVRAWPMKRLKRLKRETAGNARMWGL 541
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 507 LIVPSESLEEVARSINLKYHDLRLHSHIYVMFEQRLKEIOKE-LALFEQVKKFTLL 565
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 542 DTPEPIDGEVTDKGSINQRAVLQWRSAKYDALY 575
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 566 PQAFTMETGELTPMKLRKIIIDRYONEIDLMY 599
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
```

RESULT 7

H64041 probable long-chain-fatty-acid-CoA ligase (EC 6.2.1.3) H10002 [similarity] - Haemophilus

C:Species: Haemophilus influenzae

C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 09-Jul-2004

C:Accession: H64041

R:Heidechmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.;

R:Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.

Science 269, 496-512, 1995

A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Frazer, C.M.; Smith, H.O.; Venter,

A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.

A:Reference number: A64000; MUID:95350630; PMID:7542800

A:Accession: H64041

A:Molecule type: nucleic acid sequence not shown; translation not shown

A:Residues: 1-607 <TIGR>

A:Cross-references: UNIPROT:P44446; GB:U12686; GB:I42023; NID:g3212175; PIDN:AAC21681.1;

C:Superfamily: Synecocystis long-chain-fatty-acid-CoA ligase; acetate-CoA ligase homolo

C:Keywords: acid-thiol ligase; coenzyme A

F:63-590/Domain: acetate-CoA ligase homology <ACL>

Query Match 9.3%; Score 286; DB 2; Length 607;

Best Local Similarity 21.0%; Pred. No. 1,9e-14;

Matches 137; Conservative 115; Mismatches 251; Indels 150; Gaps 28;

```
QY 15 ILERLHMAKTRPQTCVAAARAANGEMRRISYAEFNNVPAIAQSLIPYGLSARPLLIYS 74
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 16 VHRIOQAQKTRMTMLYRK-ENGLMRDISMKRPOQLNLSALAHNIIDVDKAIIRA 74
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 75 GN-----DLEHLQAFGAMVAGIPYCVSPAYSLSDLAQRHIVGLQGLVFAAD 127
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 75 HNNERTWIVDIATLQIR-----AIVPPIYATNTAQQAEP--ILNHADVKILFVCG 122
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 128 AAFQQAIFETILPDDVPAIFTRGELAGRTVSPSLERQPGLEAD-----NAFATGE 181
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 123 QEOYDQTLR-IANHCPRK-----QKIVAKSTIQ---LQDPLCTWESFKTQS 168
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 182 DT-----IAKFLTSGSTKLPKAVPTTORMLCANQOMLQTFPVGSEEP 225
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 169 NAQODELTQRLNQKSLDLETTIYTSGTGEPKGMVLDYANLA--HQLETHDLSLAVTDQ 226
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 226 PVLVDMLPMNHTFGSGSHNIGIVLVNGGTYYLDDGKPTAQFAETLRNLSISPTAYLTV 284
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 227 DISLSPSPSHIFERAW-AAVILHRCALCYLEDTNQVBSA-----LTELIPFLMCAV 278
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 285 PKMEELVGA---LRDSTLERFPARMKLPFAAGLSQGIWD-RDQRYA----- 331
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 279 PRYERKIYAAVLDKVAKPLRQ-----IMFWAISGVQYFDLRANKALPPLKKQ 331
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 332 -----BOHCERIRMA-----AGIMETTAB--SC 354
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 332 PALADKLVLSTKRLQLGRIKQWPCGAKLERALIGFPAIGINIKLGGMETTATVSG 391
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 355 --TFTTGPLSMAGYGLPAPGCEVVLVVDGKLEGRFHPHWSGYWRAPBONAOAPDE 412
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 392 WHDFQNPNS----IGTLMPKAEVK---IGENNEILVRCGMWKKGYKKPEETQAFTED 444
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 413 GYVCSGDAIKLADPAPOKGLMFDGRIADDFKLSGVFVSVGPRLTRAIVLGGSYVLDYV 472
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 445 GFLKTDAGEF---DEQGNLFTTDRIKELMKTSNKGTYA--POYESIKIGDKFLBOJA 498
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 473 VAAPDECLGLVFPRLDLC-RALSGLGKREAS-----DAEVLASEPVRAWPMKRLR 526
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 499 IINDACKYVSALIVP---CFDSLEEVAKQINIKYHDLRLKNSIILMGFERIYAVQK 554
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 527 EATGNARIMWGLDTPPSIDKGEVTDKGSINQRAVLQWRSAKYDALYGED 579
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 555 E-LAHFEQVKFTLLSQAFSIRKLGSIYPTLKLRRKYLERYRQIEMAYHSQE 606
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
```

RESULT 8

H82069 probable long-chain-fatty-acid-CoA ligase VC2484 (imported) - Vibrio cholerae (strain NI

C:Species: Vibrio cholerae

C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004

C:Accession: H82069

R:Heideberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;

R:Chadson, D.; Ermolaeva, M.D.; Vamathevan, J.; Baas, S.; Qin, H.; Dragol, I.; Sellers, P.

Nature 406, 477-483, 2000

A:Title: DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A:Reference number: AB2035; MUID:20406833; PMID:10952301

A:Accession: H82069

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-601 <HEI>

A:Cross-references: UNIPROT:Q9KPB9; GB:AE004318; GB:AE003852; NID:g9657063; PIDN:AAF9562

A:Experimental source: serogroup O1; strain N16961; biotype El Tor

C:Genetics:

A:Gene: VC2484

A:Map position: 1

C:Superfamily: Synecocystis long-chain-fatty-acid-CoA ligase; acetate-CoA ligase homolo

RESULT 4

H87662

fcrlyl-CoA synthetase [imported] - *Caulobacter crescentus*C/Species: *Caulobacter crescentus*

C/Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Jul-2004

C/Accession: H87662

R/Merman, W.C.; Feldblum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eissen, J.; Heidelberg, J.

B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kjol

n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A/Title: Complete Genome Sequence of *Caulobacter crescentus*.

A/Reference number: A87249; MUID:21173698; PMID:11259647

A/Accession: H87662

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-621 <SNO>

A/Cross-references: UNIPROT:Q9A368; GB:A8005673; NID:g13425038; PIDN:AAK23300.1; GSPDB:C

C/Genetics:

A/Gene: CC3338

Query Match 31.5%; Score 972; DB 2; Length 621;

Best Local Similarity 38.3%; Pred. No. 1.5e-68;

Matches 219; Conservative 86; Mismatches 249; Indels 18; Gaps 6;

QY 11 PGRILRLHMAKTRPQTCVAAAPAAAN-GEWRRISVAMFHNVAIAQSLPLVGLSARP 69

DB 50 PRSIAHLIAAKAEHPDPRYLKQREPHNGPRGVGTGEAHRVAGVAGWILDKLQPHDS 109

QY 70 LLIVSGNDLHLQLAFAMTAAGTPYCPVSPAYSLISODLAKRHITVGLLOGLVFAADAA 129

DB 110 VMLISGSIHAAWMTLCAAYAGVPAAPISPAVSLISTDHAKLKCFEKVAPRVVFAQSGA 169

QY 130 PFGRAITLIPDDVPAIFGTGELAGRTVSPDSLEQPGIEADNAAAPGPTIAFLF 189

DB 170 MFGALLATLALDPSLVITTAADGSGAITFAEVAATTPPAVAADAAVGPATVAKYLF 229

QY 190 TSGSTLKPAAVPTTORMLC--ANQOMLQTFPVGGEPPVLDWLPWNHTFGSHNIGI 246

DB 230 TSGSTLKPAAVPTTORMLC--ANQOMLQTFPVGGEPPVLDWLPWNHTFGSHNIGI 287

QY 247 VLYNGGTYVDDGKPTAOGAETLRNLSPTAYLTVPKGMBELVGLERDSTLRERFF 306

DB 288 VINGGTYVDDGKPTAOGAETLRNLSPTAYLTVPKGMBELVGLERDSTLRERFF 347

QY 307 ARKKLFPFAAGSQGIWDLDRVAEQHCERIRMAAGLMTETAPSCFTTGTGLSAGY 366

DB 348 KMLRYNGYGAATLSNDYERIKQALVAEFGHRLPLTTMYGATET-QGIVTHWITERVGL 406

QY 367 IGLPAPCEVAVLVVDKLEGRFHGHVMSGYWRAPQNAQAPDEEGYCSGDAIKLADP 426

DB 407 VGLPLPGLQKLAAPSGSKYEVAVGPTVAAGYHNDPRTAAAPDEEFYGLGDAAKRVDP 466

QY 427 ADPQKGLMFGRIAEADFKLSGVSVPGLPTRAIVLEGGSYLDVVAAPDRECLGLV 486

DB 467 DDPAKGLVFGKRVEDFKLSGTVSVGLRPDLVAAACSPYIHDAVITGDKAFIGAMLV 526

QY 487 PRLLDCAALSGLGAEADAVLASSEPR--AMPADMLKRLNRATGNAARIMVGLDIT 543

DB 527 P-----SPAGLAALVADPGPEPTLEKVAIIKDKLAAFNAAAGSSRRVAFETLLTE 578

QY 544 PSIDKGEVTDKSGINORAVLQWRSKADYALY 575

DB 579 PSIDAGEITDKGIVNRCVLERADRVVALLY 610

RESULT 5

E87040

acyl-CoA synthase [imported] - *Mycobacterium leprae*C/Species: *Mycobacterium leprae*

C/Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Jul-2004

C/Accession: E87040

R/Cole, S.T.; Elgimeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho

R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Felwell, T.; Fraser, A.; Hamlin, N.; Holroyd, eam, M.A.; Rutherford, K.M.

Nature 409, 1007-1011, 2001

A/Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq

A/Title: Massive gene decay in the leprosy bacillus.

A/Reference number: A86509; MUID:21128732; PMID:11234002

A/Accession: E87040

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-476 <SNO>

A/Cross-references: UNIPROT:Q50017; GB:AL450380; NID:g13093059; PIDN:CMC31432.1; GSPDB:GT

C/Genetics:

A/Gene: xciC

C/Superfamily: 4-coumarate-CoA ligase; acetate-CoA ligase homology

Query Match 9.7%; Score 298; DB 2; Length 476;

Best Local Similarity 26.4%; Pred. No. 1.5e-15;

Matches 135; Conservative 61; Mismatches 196; Indels 120; Gaps 24;

QY 44 SYAMFHNVAIAQSLPLVGLSARPLIVSGNDLHLQLAFAMTAAGTPYCPVSPAYSL 103

DB 38 SVABRVGARGLVAVLATP--TASTVLAIT-----GGLIAGVFPVVPADIGV 82

QY 104 LSQDLAKRHITV-----GLDQGLVFAADAAAPGPTIAFLF 157

DB 83 VER-----RHMLTDSGAQAWLCPGP--SADSA-----PDGLFHI-----P 115

QY 158 VSPDSLEQPGIEADNAAAPGPTIAFLF 217

DB 116 VQIDA-----RSMRYPEPSPDATTAVVYTSITGPPKGVLSRAAIVADLDAQA 167

QY 218 FPGVGEPPVLDWLPWNHTFG-----GSHNIGIYLVNGGTYVDDGKPTAOGAETLR 271

DB 168 WQMTATD--VLVHGLPLFVHGLVGLGSLR-----GNFVHTGKPTPAVYAG-- 216

QY 272 NLSEISPTAYLTVPKGMBELVGLERDSTLRERFFARMKLFPAAGLSGINDRLDVA 331

DB 217 -CSEAGSGLYFGVPTWSSRLVA---DEAV-ARALPAPRLVSGSASLPVPVDRRLAHLT 270

QY 332 EQHGERIRMAAGLMTETAPSCFTTGTGLSAGY 384

DB 271 GHRPIER-----YGSTESLITLSTLADGERRAGWGLPLAGVQTRLVDSGGPVPYDGE 324

QY 385 LGRF--HGPVMSGYWRAPQNAQAPDEEGYCSGDAIKLADPADPQKMGFDRIAD 442

DB 325 TVRILQVRSPTMGVGLNREATAAFADEDGWTRTG--VAVVDGSGMHR---IVGREGVD 380

QY 443 FKLSGVFVSVPGLRTRAVLEGGSYLDV--VVAAPDRECLGLVPRLLDCAALSGLGKE 501

DB 381 LKILGVRIGAGRI--EMALLGHPDVREAVAVGLPD-EDIGORIVAIV-----VGAE 429

QY 502 ASDAEVLAS-----EPVRAWFMADMLKR 523

DB 430 ALDADELINVAOQLSIHKRPREVRVVDALPR 461

RESULT 6

AG0066

probable AMP-binding enzyme-family protein YPO0537 [imported] - *Yersinia pestis* (strain CC/Species: *Yersinia pestis*

C/Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Jul-2004

C/Accession: AG0066

R/Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tildall, R.W.; Holden, M.T.G.; Prentice, M.B.;

deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; i

ll, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall, f

Nature 413, 523-527, 2001

A/Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.

A/Reference number: AB0001; MUID:21470413; PMID:11586360

A/Accession: AG0066

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-601 <KUR>

A/Cross-references: UNIPROT:Q82IG5; GB:AL590842; PIDN:CAC89394.1; PID:g15978630; GSPDB:GT

Qy	424	UDDPADPOQGLMFDGIAEDFKLSSGVFSVGBLRRAYLLEGGSYLDDVVVAAPDEECGL	483
Db	425	AVADDPGRGFYFDGMAENFKLQGTGTWVAVGLRLQVLMFAGLLRDVATIGENAEELGA	484
Qy	484	LVFPRLDCLRLSLGLSKESAPDAEVLASEPVRAWFADPKLRLREATGNASRLMTWGLDT	543
Db	485	LVFPRLPLRLRELVRSQHLSDAEITLRHPSVRQIYAKLSAHOKASGSGASHRMRLVMD	544
Qy	544	PPSIDKGEVTDKGSINQRAVLQWRSAKYDALY	575
Db	545	ALREFKEGVTDKGSINQRAVLLHRKELVESLY	576

RESULT 2

long-chain-fatty-acid-CoA-ligase Atu1416 [imported] - Agrobacterium tumefaciens (strain C:/Species: Agrobacterium tumefaciens
C:/Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
C:/Accession: AB2750
R/Mood, D.W.; Setubal, J.C.; Kaul, R.; Monk, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.
erage, G.; Gillet, P.; Zhang, C.; Guenther, D.; Kutuyavin, T.; Levy, R.; Li, M.; McCell
i, Karp, P.; Romero, P.; Grant, S.
Science 294, 2317-2323, 2001
A:/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:/Reference number: AB2577; MUID:21608550; PMID:11743193
A:/Accession: AM2750
A:/Status: Preliminary
A:/Molecule type: DNA
A:/Residues: 1-611 <KUR>
A:/Cross-references: UNIPROT:O8UFI1; GB:AE006688; PIDN:AAL42422.1; PID:g17739834; GSPDB:G
C:/Experimental source: strain C58 (Dupont)
C:/Genetics:
A:/Gene: Atu1416
A:/Map position: circular chromosome

Query March 42.0% Score 1294; DB 2; Length 611;
 Best Local Similarity 46.5%; Pred. No.5.6e-94;
 Matches 266; Conservative 84; Mismatches 220; Indels 2; Gaps 2;

5 EALLPFGRIILERLEHMAKTRPEQTCVAAARAANGEMRRISYAEHFNHVRALAGSLPEYL 64
 26 DPLRPYPOKLBERLLHMCSSAPERTWMADRGREWRVSYAEALDKIRIRIGCFLLDHL 85
 65 SAERPLIVSNDENHLQAFGAMYAGIPYCPVPAYSLSDQALAKLHNITGLDPLGVF 124
 86 SVSRPLVLSNSIEHMLMTLAQHVGLASALTPRAY-TSADLTKLADRTGQITPGVVF 144
 125 AADAPFORALETLLPDDVPAIPTRGELAGR-RIVSPDSLLEPGGIEADNAFAATPDT 183
 145 AEDTTPFRBAIGEVEFDGDTPLVGLRNLPEDRSNTFNHFTLLTETPTEAVDRAFAVGPD 204
 184 IAKLFTSGSGTKLPAVPTTQRMLCANQOMLQTFPVNGEERPLVLMIPNNHFFGSHN 243
 205 VAKLEFTSGTGSPPAVIQTQRMLCNSQEMADCGIFREEPPVVDAPNNHHAAGKV 264
 244 IGIVLNGGTYYLDDGKPTAOGFAETLNLSEISPTAYLTVPKCMEEBLVGLERDSTLRE 303
 265 FNLVLVNGGTYYYIDRGKPSPAQIQGLTDLNDLADISPTVYFNVPAHEMLVQMRDEALCR 324
 304 RFPARMKLFPPAAAGLSGCIWDRLDVRAEQHCGERIRMMALGMLTETPASTPTTGPLSM 363
 325 SFPPDLKMLMAGAGMAQHTDALTELSMATVGHAVLGAAGLSSTETAPFSLFCTEPODK 384
 364 AGYIGLPAPGCEVYLVPYDGLKEGRFHGPHVMSGYWRAPBEQNAQAFBEGGYCSGDAIKL 423
 385 PGNIGIPAQGTMTKLVPFDGRYELRLKGPNTIPGWRNGELTLAAPDEEGFYRIGDTYKF 444
 424 ADPPAPQKGLMFDGRIAADFKLSSGVFVSUPLRLTRAVLBEGSYVLNVVNAAPRECEGL 483
 445 AVADDPREGGYFDGMAENFTLQGTGTVAAVPLRLQALVNMFAGLIRDAVITIGENRAEBGA 504

QY 48 LVPEPLIDCRALSGJGKEXSDAEVLAASEPVFMADMKRLNREATGNASIMVGLDT 544

DB 505 LVVPEPLIALRELVRGSGHLSDAEIIIRHSPVROQIYAKLSAHQKASSGASAKVHILMED 566

QY 544 PPSIDKGEVTYDKGSINQRAVILWOMRSKADYAL 575

DB 565 ALRPEKGEVTDKGSINQRAVILHRLREVEST 596

RESULT 3

fem10y1-CoA synthetase [imported] - *Caulobacter crescentus*
 C.Species: *Caulobacter crescentus*
 C.Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Jul-2004
 C.Accession: F87297
 R.Nieman, W.C.; Feldblum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.F.
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolonin
 N. J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
 A.Title: Complete Genome Sequence of *Caulobacter crescentus*.
 A.Reference number: AB7249; MUID:21173698; PMID:11259647
 A.Accession: F87297
 A.Status: preliminary
 A.Molecule type: DNA
 A.Residues: 1-596 <STO>
 A.Cross-references: UNIPROT:Q9AB43; GB:AE005673; NID:G13421550; PIDD:AAK22378.1; GSPDB:GN
 C.Genetics:
 A.Gene: CC0391

Query Match	32.5%	Score 1002.5	DB 2	Length 596
Best Local Similarity	40.5%	Pred. No. 5.4e-71		
Matches 227	Conservative 81	Mismatches 227	Indels 25	Gaps 7
QY	18	LEHNAKTPPEQTCVBARANGEMWRISYAEFHNHRAIQSLPYGSAERPLIVSGND	77	
DB	49	LARMAAQAQPDWRVWLEARDQEG--WRTVTYADAKAQVETALGGTGLGRGCPILLIILRANG	107	
QY	78	LEHLOLAGAMATAGIPYCPVSPAYSLISODLAKRITVGLLOGLVPAADAAFPQRIET	137	
DB	108	VDHARIAAAMSLGAPAAVPSFOYGLKADITRLQHAVERLKRAAVVADDAEAAGDILAA	167	
QY	138	ILPDDVPPIFTREGELAGRRTVSPDSLLEOPGIEADNAFAATPDTIAKEFLTSGTKLP	197	
DB	168	PPLAGLPVVVSUNARPG--DVAFEALLKSP-----PLAPQAPDDVAKLLISGSGTKP	219	
QY	198	KAVPTTQRMLCANQOMLQTFPVFGSEPPVLYDMLPWNHTFGSGHNIQVLYNGCTYYLD	257	
DB	220	KAVVCTHNIATINAAQIEACVA--DEDPVLVVASAPWSHSGANAIIMHWLHRCGTLYID	277	
QY	258	DGKPTAAGPAETLRNLSEISPTAYLLTVPRKGMELVQALREDSLTRERFFARMKLPFPAAL	317	
DB	278	AGQVPBGRGEIVRNLISEVAITTYHNKVPAGMAMFVBLEDEDLAARFFEKVRVLQYGA	337	
QY	318	GLSGQIWMRLDVAEOCHGERIRMAAGLQMTETAPSCFTPTGPLSMAGYIGLPAPGCEVK	377	
DB	338	SMAGSILDRVQAVAAVRTGERTITPAAGYGATGTPPACNIMHMLNARGMVGLPFGTAIVK	397	
QY	378	LVPV--DGKLEGRFHGPHVMSGYWRAPEDONAAQFDEBGYYCSGDALITLADPADPQKGLMF	435	
DB	398	LVPAGEGGFETLRVKGPOVSPGYLDQPEATAQAFDEBGFRLDDAALAPEDDAAGIVF	457	
QY	436	DGRIADFPKLSGCVFVSGPLRTBRAVLLEGSGYLDVVVAAPRDEKCLGLVFPRLDLCRAL	495	
DB	458	DGRLVENFRLAAGTITPAAGALRAVAVSAIGVSDAVVCGEGGEGVGMLEF--LDKAV	514	
QY	496	SGLGKEASDAEVLAASEPVPRAMPADWLKRLINREATGNASRIIMVGLDTPPSIDKGEVTDK	555	
DB	515	ERLQ-----SPEAVRAAIVEALSRMMAAKGGGGRIRARALILDGAPDAASGELTDDK	565	
QY	556	GSINQRAVLQWMSAKVDLY	575	
DB	566	GTINQALARDRRPKELERLP	585	

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2005, 23:20:35 ; Search time 39 Seconds
(without alignments)
1453.120 Million cell updates/sec

Title: US-09-750-986D-30

Perfect score: 3082

Sequence: 1 MRSLEALPPGRILRLERH.....KVDALYRGDSMLRDEATL 589

Scoring table: BLOSUM62
Gapop 10.0 , Gapept 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1294	42.0	591	2	feruloyl-CoA synth
2	1294	42.0	611	2	long-chain-fatty-a
3	1002.5	32.5	596	2	feruloyl-CoA synth
4	972	31.5	621	2	feruloyl-CoA synth
5	298	9.7	476	2	acyl-CoA synthase
6	289	9.4	601	2	probable AMP-bind
7	286	9.3	607	2	probable long-chain
8	283.5	9.2	601	2	probable long-chain
9	273	8.9	677	2	probable long-chain
10	270	8.8	721	2	lipidosis - mouse
11	265	8.6	601	2	probable long-chain
12	261	8.5	536	2	2,3-dihydroxybenzo
13	260	8.4	536	2	2,3-dihydroxybenzo
14	258	8.4	566	2	long-chain-fatty-a
15	257.5	8.4	600	2	probable fadD15 -
16	257	8.3	508	2	probable long chain
17	256	8.3	536	1	2,3-dihydroxybenzo
18	255.5	8.2	687	2	hypothetical prote
19	254	8.2	656	2	probable long-chain
20	251	8.1	558	2	probable acetyl-co
21	250.5	8.1	683	2	hypothetical prote
22	250	8.1	473	2	probable fadD36 pr
23	247	8.0	645	2	probable long-chain
24	246	8.0	699	2	long-chain-fatty-a
25	244	7.9	638	1	long-chain-fatty-a
26	244	7.9	4342	2	probable non-ribos
27	243	7.9	722	2	brain-specific lon
28	242.5	7.9	600	2	probable long-chain
29	242	7.9	509	2	probable acid-CoA

30	242	7.9	647	2	T30892	hypothetical prote
31	240.5	7.8	550	1	A26772	Photinus-luciferin
32	239.5	7.8	598	2	T35513	probable long-chain
33	238.5	7.7	696	2	S74984	long-chain-fatty-a
34	237	7.7	547	2	A70551	probable acid-CoA
35	236	7.7	827	2	S72741	probable polyketid
36	236	7.7	1188	2	D86969	probable acyl-CoA
37	235.5	7.6	530	2	B87369	probable long-chain
38	235.5	7.6	612	2	T35430	probable long-chain
39	235.5	7.6	720	2	E84456	probable acyl-CoA
40	235	7.6	584	2	C75364	probable long-chain
41	233	7.6	525	2	AG0232	Yersiniaabactin sid
42	233	7.6	525	2	T17443	salicyl-AMP ligase
43	232.5	7.5	823	2	D70386	probable polyketid
44	232	7.5	555	2	C83444	probable AMP-bind
45	232	7.5	650	2	H84356	long-chain fatty-a

ALIGNMENTS

RESULT 1

G97531
feruloyl-CoA synthetase (A238746) [imported] - Agrobacterium tumefaciens (strain C58, C
C/Species: Agrobacterium tumefaciens
C/Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 09-Jul-2004
C/Accession: G97531
R/Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quirollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tume
A/Reference number: A97359; MUID:21608551; PMID:11743194
A/Accession: G97531
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-591 <KIR>
A/Cross-references: UNIPROT:Q8UFI1; GB:AE007869; P1DN:AAK87208.1; P1D:G15156488; GSPDB:GT
C/Genetics:
A/Gene: AGR_C_2614
A/Map position: circular chromosome

Query Match	Score	DB 2;	Length	591;
Best Local Similarity	46.5%;	Pred. No. 5.4e-94;		
Matches 266;	Conservative 84;	Mismatches 220;	Indels 2;	Gaps 2;
QY	5	EAALPPGRILRLERHMAKTRPEOTCVARARANGEMRISVAFEMFNRAIAGSLPYGL	64	
DB	6	DPLGPIYQKLNRLHMKRSAPERTWADROGRBPRRVYAEALDKIRIGPLDHD	65	
QY	65	SAERPLIVSGNDLEHLQIAFGAMYAGIPYCPVPAYSLSDILAKLRHIVGLDPLVF	124	
DB	66	SVSRPLIVSSENGIEHLAVLAQHVGIASAITPAVA-TSADUTLKADIRGQITPMVF	124	
QY	125	AADAAFPQRAIEITLPDVAITFRGELAGR-RTVSDSLLEORGGIEADNAPATPDT	183	
DB	125	AEATATPRRALGCFDGTPLVGLRLNPEDRSNTFHETLTLEPTBAVRAAPAVPDT	184	
QY	184	IAKFLPFGSGTKLPKAVPTQRMICANQOMLTQTPVFGSEPPVLDMLPMNTHFGSSHN	243	
DB	185	VAKFLPFGSGTKPKAVIQTRMICSQEMTADICGTFREBPVVDMAAMNHTAAGNKV	244	
QY	244	IGILVNGGTYYLDDGKPTAGPAETLRNLSEISPTAYLITVPKWEELVGALESDSTIRE	303	
DB	245	FNILVYNGGTYYLDRGKSPQAQIGQLTDNLNDISPTWYFNVAVGHEMLVQAMKDEALCR	304	
QY	304	RFPARMKLFFPAAAGSOGIWDRLDVAEQHCERIRMAAGLMTETPAESCTPTTGLSM	363	
DB	305	SFPRDLKMLMYAGAGVAQHTMDALTELSTAVGHAVALMGSLGSLSTETAPSLCTBEQDX	364	
QY	364	AGYIGLPAPGCEVKLVVDGKLEGRPHGPHVMSGWYAPQONQAPDEEGYSGDAIKL	423	
DB	365	PGNITGLPAGVMTKLVDFDGRYEURLKGPNTITGWMNGSLTAADDEEGFYRIQDITVKF	424	

QY 470 AspValValAlaAlaProAspArgLysCysLeuGlnLeuValPheProArgLeu 489
Db 1420 CAATATATGTTGGTGAAGGAGAAATTCCTTCTGCGATTGTGGCGCTGTT 1479
QY 490 LeuAspCysArgAla-----LeuSerGlyLeuGlnLysGlnAlaSerAspAlaGlu 506
Db 1480 GCAGATGTAAAGAGCTGGTGTGAGATCAAGCAAT---CCGATATACGACAAAGAA 1536
QY 507 ValLeuAlaSerGluProValArgAlaTrpPheAlaSerTrpLeuLysArgLeuAspArg 526
Db 1537 ATGGTTTGCATGACGACGATGATGAGAAATTCAGAAAGAGTTTGACGAATATAT--- 1593
QY 527 GluAlaThrGlyAsnAlaSerArgLysMetTrpValGlyLeuAspTrpProSer 546
Db 1594 GAAACCTTGCACAGTATGAACGTTGAAGAAATTCATTTAATGCCAAGCCATTACG 1653
QY 547 IleAspLysGlyGlnValIleThrAspLysGlySerIleAsnGlnArgAlaValLeuGlnTrp 566
Db 1654 ATAGAGAGCGAGAAATGACACCGACATTCGAAATTTGAACGTAAAGATCATTTATGCA 1713
QY 567 ArgSerAlaLysValAspAlaLeuTyr 575
Db 1714 TGCCAGGCGCAGATAGAGAAATGTAT 1740

RESULT 15
US-10-470-565-1
; Sequence 1, Application US/10470565
; Publication No. US20040126870A1
; GENERAL INFORMATION:
; APPLICANT: Societe des Produits Nestle S.A.
; TITLE OF INVENTION: NCC2705 - the genome of a Bifidobacterium
; FILE REFERENCE: 80290/WO
; CURRENT APPLICATION NUMBER: US/10/470,565
; PRIOR FILING DATE: 2003-07-29
; PRIOR APPLICATION NUMBER: EP 01102050.0
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 2256646
; TYPE: DNA
; ORGANISM: Bifidobacterium longum
US-10-470-565-1

Alignment Scores:
Pred. No.: 3 77e-20 length: 2256646
Score: 314.50 Matches: 146
Percent Similarity: 38.70% Conservative: 87
Best Local Similarity: 24.25% Mismatches: 235
Query Match: 10.20% Indels: 134
DB: 17 Gaps: 23

US-09-750-986d-30 (1-589) x US-10-470-565-1 (1-2256646)

QY 14 IleLeuGlnArgLeuGlnHisTrpAlaLysThrArgProGlnGlnThrCysValAlaAla 33
Db 354867 CTTACCGATCGCCTT-----GCCCGACCGCGCGACGACGGAATCGCGCCAG 354917

QY 34 ArgAlaAlaAsnGlnLysTrpArgArgIleSerTrpAlaGlnMetPheHisAsnValArg 53
Db 354918 AAAATCGCGCCCGCGCTGCGCAAGCTCACCGCGCGAATTTCCACGAGCGGTGTC 354977

QY 54 AlaIleAlaGlnSerLeuLeuProTyrGlyLeuSerAlaGlnLysArgProLeuLeuVal 73
Db 354978 TCGCGCCGCAAGGCGCTGATCGCGCTGCGCATCCCAAGGCGACGCGGTGACCATTTT 355037

QY 74 SerGlyAsnAspLeuGlnHisLeuGlnLeuAlaPheGlyAlaMetTrpAlaGlyIlePro 93
Db 355038 TCCCTCCACCGCGTGAAGATGGGGCATTTCTGATTTGCCACGCGCGCGGTGGGCGGTA 355097

QY 94 TyrCysProVal-----SerProAlaTyrSerIleLeuSerGlnAsp 107
Db 355098 AGCGTGCCGATCTATGACACCGACTCGCCCGCGAGCGCCAGCGCATGAAAGATTCG 355157

QY 108 LeuAlaLysLeuArgHisIleValGlyLeuLeuGlnProGlyLeuValPheAlaAlaAsp 127
Db 355158 GTGGTCAAG-----CTGGCCTTTGCCGACAT 355184

QY 128 AlaAlaProPheGlnArg-----AlaIleGlnTrpIle 138
Db 355185 CGCGAGCGCTTCACCGCGCTCGACTCGTCAGACACCTGCCCTGTGAAGCAAT 355244

QY 139 LeuProAspAspValProAlaIlePheThrArgGlyGlnLeuAlaGlyArg-----Thr 157
Db 355245 CTCATGATCGAGGCGCAACGCGCTG-----GGCGCTTGAAGCGCTAGCGGTGACC 355295

QY 158 ValSerPheAspSerLeuGlnGlnProGlyGlyIleGlnAlaAspAsnAlaPheAla 177
Db 355296 GTTTCCGACGAGAACTGCACGAGCGCGGTGCCACCGCTGCCACGCGAC----- 355346

QY 178 AlaThrGlyProAspThrIleAlaLysPheLeuPheThrSerGlySerThrLysLeuPro 197
Db 355347 -----CTGGCCACCATGCTCTACCTTCGGGCTTCACCGCGAC----- 355388

QY 198 LysAlaValProThrThrGlnArg-----MetLeuCysAlaAsnGln 212
Db 355389 AAGGTGCGCACTGACCCACAAAGAACTTGTCTCATCACTTCGCTGCGACGCG 355448

QY 213 MetLeuLeuGlnThrPheProValPheGlyGlnGluProProValLeuValAspTrpLeu 232
Db 355449 TTG-----CATGAAGTGGTCTTGAAGACCATCGCGCGCTGCTTCTG 355496

QY 233 ProTrpAsnHisThrPheGlyGlySerHisAsnIleGlyIleValLeuTyrAsnGly 252
Db 355497 CGCTGCGCCACTGCTGCGCGATTCATCCAGTACGCTTCCTCCGATGACGGC 355556

QY 253 ThrTyr-----TyrLeuAspAspGlyLysProThrAlaGlnGlyPheAlaGlnThrLeuArg 271
Db 355557 GTGGTGGTTATCTGCGGACACCA-----ACCTGCTGCGC 355595

QY 272 AsnLeuSerGlnLysSerProThrAlaTyrLeuThrValProLysGlyTrpGlnGlnLeu 291
Db 355596 GACCTGTTGTCATTGAACCGACCTACCTGCTGGGCGTCCGCGGCTTCGAGAAAGTA 355655

QY 292 ValGlyAlaLeuGlnArgAspSerThrLeuArgGlnArgPheAlaArgMetLysLeu 311
Db 355656 TACACGCGCCCTCGACACGACGCGGT-----GCCGCTGAAAGGGC 355697

QY 312 PhePhePheAlaAlaAlaGlyLeuSerGlnGlyIleTrpAspArgLeuAspArgValAla 331
Db 355698 CGCCTGTTGTCAAAGCGCCGAGGCGCGCGCTGTGAGCCGCAAGGACGAGCGCGC 355757

QY 332 GluGlnHis----- 334
Db 355758 GAGCAGACACGTTTGCGCAAAATCGCGAGCGCGCAATGCAAAACGCTGCTACCGC 355817

QY 335 -----CysGlyGlnArg----- 338
Db 355818 ACGGTGCGGTCGTCTGCGCCGAAAGATCAATAGTAGTGGCTGCGCGCGACCACTG 355877

QY 339 -----IleArgMetAlaGlyLeuGly 346
Db 355878 TCCCTGATCTGGCGCATTTCTAACAAGCGCATGCACTGCCAGATTCAGGGTTATGGC 355937

QY 347 MetThrGlnThrAlaProSerCysThrPheThrThrGlyProLeuSerMetAlaGlyTyr 366
Db 355938 ATGACGAAACCGCCCGCGCTTCCGCGCGCGGTGACGCAATGATGATCGGACCC 355997

QY 367 IleGlyLeuProAlaProGlyCysGlnValLysLeuValProValAspGlyLysLeuGln 386
Db 355998 GTAGGCGAGCGCGCGCGCTCTCGATTCGC---ATTCCACGAAAGCGAACTGCGAG 356054

QY 387 GlyArgPheHisGlyProHisValMetSerGlyTyrTrpArgAlaProGlnGlnAsnAla 406
Db 356055 GTCAAG-----GCGCGAAAGTTCCTCGCGGTACCACAATCTCGCGAGAAAGCGCC 356108

Qy 528 aThrGlyAsnAlaSerArg-----IleMetTrpValGlyLeuLeuAspThrProAsp 545
 Db 6728952 ----GGTGGCTCCAGGCTGCGAGACCACTCAGAACTCCCTCCGCGCGACCT 6729007
 Qy 545 oSerIleAspLysGlyValThrAspLysSerIleAsnGlnArgAlaValLeuG1 565
 Db 6729008 CGACATCGAACACGCGAGATGATCCCGAGCTCAACAGCTGAGCGCGCTCGCAACG 6729067
 Qy 565 nTrpArgSerAlaLysValAspAlaLeuTyArgGlyValAspGlnSer 581
 Db 6729068 CGAGTACACAGCATCTGATCGACGATGATCGCGGGTCCGCGAGGCT 6729116
 RESULT 14
 US-10-369-493-34073 Application US/10369493
 / Sequence 34073.
 / Publication No. US20030233675A1
 / GENERAL INFORMATION:
 / APPLICANT: Cao, Yongwei
 / APPLICANT: Hinkle, Gregory J.
 / APPLICANT: Slater, Steven C.
 / APPLICANT: Goldman, Barry S.
 / APPLICANT: Chen, Xianfeng
 / TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 / FILE REFERENCE: 38-10(52052)B
 / CURRENT APPLICATION NUMBER: US/10/369,493
 / PRIOR FILING DATE: 2003-02-28
 / PRIOR APPLICATION NUMBER: US 60/360,039
 / NUMBER OF SEQ ID NOS: 47374
 / SEQ ID NO 34073
 / LENGTH: 1746
 / TYPE: DNA
 / ORGANISM: Cytophaga hutchinsonii
 US-10-369-493-34073
 Alignment Scores:
 Pred. No.: 1,4e-25 Length: 1746
 Score: 324.00 Matches: 144
 Percent Similarity: 40.23% Conservative: 101
 Best Local Similarity: 23.65% Mismatches: 264
 Query Match: 15 Indels: 100
 Gaps: 26
 US-09-750-986d-30 (1-589) x US-10-369-493-34073 (1-1746)
 Qy 26 ProGlnGlnThrCysValAlaAlaArgAlaAlaAsnGlyLutTrpArgArgIleSerTyr 45
 Db 37 CCCAAACAGATGCACTGCGCAACCAA--GAGAAATGCGACCTGATTAATTCCTTCT 93
 Qy 46 AlaGluMetPheIleAsnValArgAlaIleAlaGlnSerLeuLeuProTyrGlyLeuSer 65
 Db 94 GTCCGCTTCAGAAATTATGATTAAGTAAGTATGCTTTTAAATAATCGGCTTTACAA 153
 Qy 66 AlaGluArgProLeuLeuIleValSerGlyAsnAspLeuGluHisLeuGlnLeuAlaPhe 85
 Db 154 AAAGACGATTAAGTACCATTTATTCACCAACCGCTCCGAAATGGAATTTATGATCTG 213
 Qy 86 GlyAlaMetTyrAlaGlyTyrProTyrCysProValSerProAlaTyrSerLeuLeuSer 105
 Db 214 GGACTCCACACAGATAGAGCGGTGAGTGTACCGATTTATCCG-----ACACTTACGGTG 267
 Qy 106 GlnAspLeuAlaLysLeuArgHisIleValGlyLeuLeuGlnProGlyLeuValPheAla 125
 Db 268 GAAGAT-----TACCGTATATTTTGTATGATGACCGCGAAGTTAAATGTTTGTG 318
 Qy 126 AlaAspAlaAlaProPheGlnArgAlaIleGluThrIleLeuProAspAspVal----Pro 144
 Db 319 TATGACGACAAATATTTTGCAGAAAGCAAGAGCGGTAAAGCGTTCCAGAACAAATCCT 378
 Qy 145 AlaIlePheThrArgGlyLeuAlaGlyArgArgThrValSerPheAspSerLeuLeu 164
 Db 379 GTAAATATATACCTTGACGATGTGAACGAGCAAAAGCTGGAAGATATACGCGCACTG 428

Qy 165 Glu-----GlnProGlyGlyIleGluAlaAspAsnAlaPheAlaIleThrGlyProAsp 182
 Db 439 GCAGACGGCCACCGGCTTCGATG---CTGATCCATATTAAGCAATATTAACGAC 495
 Qy 183 ThrIleAlaLysPheLeuPheThrSerGlySerThrLysLeuProLysAlaValProThr 202
 Db 496 GATCTGCTTACCTTATTAATTAAGTCCGGAACCTACAGTAAGCCCAAGGCTGTATGCTT 555
 Qy 203 ThrGlnArgMetLeuCysAlaAsnGlnGlnMetLeuLeuGlnThrPheProValPheGly 222
 Db 556 ACACATACAAATATTAATTAAGTATTAATTCGACGACGATTCGCTGATGCGGTT----- 609
 Qy 223 GlnGluProProValLeuValAspTrpLeuProTyrAsnHisIleThrPheGlyGlySerHis 242
 Db 610 AATCACAAATCATTAAGCACTTAAGTTTATTCCTGTGCGCACGCTGAT-----GAA 660
 Qy 243 AsnIleGlyIleValLeuTyr-----AsnGlyGlyThrTyrTrpLeuAspAspGlyLys 260
 Db 661 CGCATGTGTGTTATCTGATACCTGCTCAGGGGGTATCTATATATGACAGAAAGTACA 720
 Qy 261 ProThrAlaGlnGlyPheAlaGluThrLeuArgAsnLeuSerGluIleSerProThrAla 280
 Db 721 GATACAGTTGCG-----GAAATTTAAAGAAAGTACGACCGCGACATG 762
 Qy 281 TyrLeuThrValProLysGlyTyrGlu-----LeuSerGlnArg 289
 Db 763 TTGTTCAGGTGCGCAAGCTGCTTGAAAAAGTATACGATTAATTTGCTTAAGGCGCT 822
 Qy 290 GlnLeuValGly-----AlaLeuGlnArg 297
 Db 823 GAATTCACGGGCAATTAATAATTCCTCTATATGGGCACTGAGCGGGTTAAATAT 882
 Qy 298 AspSerThrLeuArgGluArgPhePheAlaArgMetLysLeu-----Phe 312
 Db 883 GATATGACAGAAATCAAGGCTTCTTTTAAATTCATCACTGAAGCTTCCATTAATAATA 942
 Qy 313 PhePheAla-----AlaAlaGly-----LeuSerGlnGly 322
 Db 943 ATCTTACGTAATATGCGAGAAAGCATTTGGGTGGAAATGATATACCTGTTTACGCGGT 1002
 Qy 323 ----IleTrpAspArgLeuAspArgValAlaGlnGlnHisCysGlyGlyLysArg 340
 Db 1003 GCTGCTTACAGCTGTTTGGCGCGT-----TTCGCGCGCGCAGATTCGG 1053
 Qy 341 MetMetAlaGlyLeuGlyMetThrGluThrAlaProSerCysThrPhe----- 356
 Db 1054 GTGATGAAAGTTTACGTTTAAACAGAGACGTGCGCGGTATTCGCTTAACAGATTCAT 1113
 Qy 357 -----ThrThrGlyProLeuSerMetAlaGlyTyrIleGlyLeu 369
 Db 1114 CCGAGTAAACACGATCACTGATCTGTGGTCCGCT----- 1149
 Qy 370 ProAlaProGlyCysGluValLysLeuValProValAspGlyLysLeuGlnGlyArgPhe 389
 Db 1150 ----TTGCGCGAGTAAAGATGAAGATCGCGGAA--GACGGAAGAAATCCGACGCC-- 1200
 Qy 390 HisGlyProHisValMetSerGlyTyrTrpArgAlaProGlnGlnAsnAlaGlnAlaPhe 409
 Db 1201 --GACCCGCAATATTAAGAAAGTACTCAAAACGCGCCGAAATTTAAACGCGCAAGCAATT 1257
 Qy 410 AspGluGluGlyTyrTyrCysSerGlyAspAlaIleLeuLeuAlaAspProAlaAspPro 429
 Db 1258 GACAGTGAAGGCTGTTTACACAGAGATATCCGCTCACTTGTATGAA----- 1305
 Qy 430 GlnLysGlyLeuMetPheAspGlyArgIleAlaGluAspPheLysLeuSerGlyVal 449
 Db 1306 AATTAATATTTAAAGATCACGACCGGTAAAGAAATGTTTAAACCTCGGCTGTAAG 1365
 Qy 450 PheValSerValGlyProLeuArgThrArgAlaValLeuGlnGlyGlySerTyrValLeu 469
 Db 1366 TACATTGCA-----CCGACGCTGTTGAATAATTAATTAAGAGTCCGCTACATTGAA 1419

```

1 CURRENT FILING DATE: 2002-05-29
2 PRIOR APPLICATION NUMBER: JP 2001-204089
3 PRIOR FILING DATE: 2001-05-30
4 PRIOR APPLICATION NUMBER: JP 2001-272697
5 PRIOR FILING DATE: 2001-08-02
6 NUMBER OF SEQ ID NOS: 15109
7 SEQ ID NO 1
8 LENGTH: 9025608
9 TYPE: DNA
10 ORGANISM: Streptomyces avermitilis
11 FEATURE:
12 NAME/key: misc feature
13 LOCATION: (4187715)
14 OTHER INFORMATION: a, t, c, g, other or unknown
15 OS-10-156-761-1

```

Alignment Scores:	
Pred. No.:	1.2e-21
Score:	335.50
Percent Similarity:	39.12%
Best Local Similarity:	23.74%
Query Matchn:	10.6%
DB:	15
US-09-750-986D-30 (1-589) x US-10-156-761-1 (1-9025608)	
Length:	9025608
Matches:	156
Conservative:	101
Mismatches:	266
Indels:	134
Gaps:	26


```

Db      ||||| ||||| : : : |||||
93  CTACCGCTACCGGTAACCTCCCGCA-----GCTCCGCGCTTCGCCCGA 137
Qy      ||||| ||||| : : : |||||
27  uGlnThrCysValAlaAlaArgAlaAlaAnglyGluTyrArgArgIleSerTyrAlaI 47
Db      ||||| ||||| : : : |||||
138  GCAGGAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 197
Qy      ||||| ||||| : : : |||||
47  uMetPheHisAlaValArgAlaIleAlaGlnSerLeuLeuProTyrGlyLeuSerAlaI 67
Db      ||||| ||||| : : : |||||
198  GCGCGCGCGCAAGGAGTCTTCGCGGTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 257
Qy      ||||| ||||| : : : |||||
67  uArgProLeuLeuIleValSerGlyLeuAspLeuGlnHisLeuGlnLeuAlaPheGlyAl 87
Db      ||||| ||||| : : : |||||
258  GCAGCGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 317
Qy      ||||| ||||| : : : |||||
87  aMetTyrAlaGlyIleProTyrCysProValSerProAlaTyrSerLeuLeuSerGlnAs 107
Db      ||||| ||||| : : : |||||
318  CATGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 362
Qy      ||||| ||||| : : : |||||
107  pLeuAlaLeuLeuArgHisIleValGlyLeuLeuGlnProGlyLeuValPheAlaAlaAs 127
Db      ||||| ||||| : : : |||||
363  CGCGGAGAGTCCGCGCTTCATCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 422
Qy      ||||| ||||| : : : |||||
127  pAlaAlaProPheGlnArgAlaIleGlu-----136
423  CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 482
Qy      ||||| ||||| : : : |||||
137  ---ThrIleLeuProAspArgValProAla-----IlePheThrArgGlyLeu 152
Db      ||||| ||||| : : : |||||
483  CGTGTGATGAGACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 542
Qy      ||||| ||||| : : : |||||
152  uAlaGlyArgArgThrValSerPhe-----AspSerLeuLeuGlnProGly 168
Db      ||||| ||||| : : : |||||
543  GAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 602
Qy      ||||| ||||| : : : |||||
168  yGlyIleGluAlaAspArgAlaPheAlaAlaThrGlyProAspThrIleAlaAsPheLe 188
Db      ||||| ||||| : : : |||||
603  CGCGATCACTCCGCGACAG-----CTCGCGCACCCCTCAT 635
Qy      ||||| ||||| : : : |||||
188  uPheThrSerGlySerThrIleLeuProGlyAlaValProThrThrGlnArgMetLeuCy 208
Db      ||||| ||||| : : : |||||
636  CTACACCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 683
Qy      ||||| ||||| : : : |||||
208  sAlaAsnGlnGlnMetLeuLeuGlnThrPheProValPheGly-----GluGluProPr 226
Db      ||||| ||||| : : : |||||
684  CGACAACTGCTCGTACATGCGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 743
Qy      ||||| ||||| : : : |||||
226  oValLeuValAspTyrLeuProTyrPheAsnHisThrPheGlyGlySerHisAsnIleGlyI 246
Db      ||||| ||||| : : : |||||
744  CGTGCAGTACTGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 766
Qy      ||||| ||||| : : : |||||
246  eValLeuTyrAsnGly-----GlyThrTyrTyrLeuAspArgGlyLeuProTh 262
Db      ||||| ||||| : : : |||||
787  -GTGCTCACCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 840
Qy      ||||| ||||| : : : |||||
262  rAlaGlnGlyPheAlaGluThrLeuArgAsnLeuSerGlnIleSerProThrAlaTyrLe 282
Db      ||||| ||||| : : : |||||
841  -----GTGCACAAAGATCATCGAAGTCTGCGCGGTGCGACGCGACATGCGC 890
Qy      ||||| ||||| : : : |||||
282  uThrValProLeuGlyTyrGluGluLeuValGlyAlaLeuGlnuArgAspSerThrLeuAs 302
Db      ||||| ||||| : : : |||||
891  TGGCGTACCGCGCGATCTTCGAGAAAGCTTCACAAAGCGGTGCGCT-----GCCAAAGGCGG 944
Qy      ||||| ||||| : : : |||||
302  gGluArgPhePheAlaArgMetLeuLeuPhePhePheAlaAla-----316
Db      ||||| ||||| : : : |||||
945  GCGCGGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1004
Qy      ||||| ||||| : : : |||||
317  -AlaGlyLeuSerGlnGlyIleTyrPheArgLeuAsp-----328
Db      ||||| ||||| : : : |||||
1005  CGCCAAAGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1064
Qy      ||||| ||||| : : : |||||
329  -----ArgValAlaGluGln-----333

```

```

Db      ||||| ||||| : : : |||||
1065  CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1124
Qy      ||||| ||||| : : : |||||
333  -----333
Db      ||||| ||||| : : : |||||
1125  CGCGCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1184
Qy      ||||| ||||| : : : |||||
334  -HisCysGlyGluArgIleArgMetMetAlaGlyLeuGlyMetThrGluThrAlaProSe 353
Db      ||||| ||||| : : : |||||
1185  CTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1244
Qy      ||||| ||||| : : : |||||
353  rCyThrPheThrThrGlyProLeuSerMetAlaGlyTyrIleGlyLeuProAlaProGly 373
Db      ||||| ||||| : : : |||||
1245  CTCTTCGTGAAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1304
Qy      ||||| ||||| : : : |||||
373  yCysGluValLeuValProValAspGlyLeuGluGluArgPheHisGlyProHis 393
Db      ||||| ||||| : : : |||||
1305  CACCGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1355
Qy      ||||| ||||| : : : |||||
393  sValMetSerGlyTyrTyrArgAlaProGluGlnAlaAlaPheAspGluGluGly 413
Db      ||||| ||||| : : : |||||
1356  CATCATGAGGCGGTACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1415
Qy      ||||| ||||| : : : |||||
413  yTyrTyrCysSerGlyAspAlaIleLeuLeuAlaAspProAlaAspProGluLeuGlyLe 433
Db      ||||| ||||| : : : |||||
1416  CTGATTCCACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1463
Qy      ||||| ||||| : : : |||||
433  uMetPheAspGlyArgIleAlaGluAspPheLeuSerSerSerGlyValPheValSerVa 453
Db      ||||| ||||| : : : |||||
1464  GCGCATCACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1521
Qy      ||||| ||||| : : : |||||
453  lGlyProLeuArgThrArgAlaValLeuGluGly---GlySerTyrValLeuAspValVa 472
Db      ||||| ||||| : : : |||||
1522  -----CCCGCGAGTGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1577
Qy      ||||| ||||| : : : |||||
472  lValAlaAlaProAspArgGlu---CysLeuGlyLeuLeuValPhe-----ProArgLe 489
Db      ||||| ||||| : : : |||||
1578  GGTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1637
Qy      ||||| ||||| : : : |||||
489  uLeuAspCysArgArgAlaLeuSerGlyLeu---GlyLeuGluAlaSerAspAlaGluValLe 508
Db      ||||| ||||| : : : |||||
1638  CTTGAGTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1691
Qy      ||||| ||||| : : : |||||
508  uAlaSerGluProValArgAlaTyrPheAlaAspTyrLeuLeuArgGlyGluAl 528
Db      ||||| ||||| : : : |||||
1692  CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1743
Qy      ||||| ||||| : : : |||||
528  aThrGlyAsnAlaSerArg-----IleMetTyrValGlyLeuLeuAspThrProPr 545
Db      ||||| ||||| : : : |||||
1744  ---GGTGGCTTCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1799
Qy      ||||| ||||| : : : |||||
545  oSerIleAspGlyGlyValAlaThrAspGlySerIleAsnGlnArgAlaValLeuGly 565
Db      ||||| ||||| : : : |||||
1800  CGACATCGAACACGCGGAGATCACCGCGAGCTCAAGCTGAAGCGCGCGGTCTCGAACG 1859
Qy      ||||| ||||| : : : |||||
565  nTyrArgSerAlaLeuValAspAlaLeuTyrArgGlyGluAspGlnSer 581
Db      ||||| ||||| : : : |||||
1860  CGAGTCAAGCATCTGATCGACGAGATGTACCGGCGGTGCGCGGAGGCT 1908

```

```

RESULT 13
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OHTA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761

```

```

Qy 101 TyrSerLeuSerGlnAspLeuAlaLysLeuArgHisIleValGlyLeuLeuGlnPro 120
Db 253 TAGCCCAATTATTTCTGAGAGATG-----ATGACTTACTGTCTGTCTATGGCAAAACC 306
Qy 121 GlyLeuValPheAla----- 125
Db 307 GCATTGCTCATCGAAGAAAGTATCATCTTCACAGACACTGCGACAGATACCCCGTGG 366
Qy 126 -----AlaAspAlaAlaProPhe-----GlnArgAla 134
Db 367 CAGGTATGACACTCGGACAGCGCGGACCTTCCCTGATTTTCCGCTATTTGACAGGAAA 426
Qy 135 ILeGluThrIleLeuProAspAspValProAlaIlePheThrArgGlyGluLeuAlaGly 154
Db 427 ACTGACCTGATCAGCCCGGATTAAGCGTCACTAGTACAGTATCGGAGAG--GCCTTA 483
Qy 155 ArgArgThrValSerPheAspSerLeuLeuGlnGlnProGlyGlyIleGluAlaAspAsn 174
Db 484 CTCAGGCGCTATTACAGACAGATTTCACAGCCACGCGAG-----GCAGAACAG 528
Qy 175 AlaPheAlaIaIaThrGlyProAspThrIleAlaLysPheLeuPheThrSerGlySerThr 194
Db 529 GCATTCGT-----TTGATGTTTACTCTGCTCCAGT 561
Qy 195 LysLeuProLysAlaValProThrThrGlnArgMetLeuCysAlaAsnGlnMetLeu 214
Db 562 GGTAAACCGAAGGTGTGTTGCTCAGTCAGAACAAATTT---CTGTCACAGCAGCAGGCT 618
Qy 215 LeuGlnThrPheProValPheGlyGluGlnProProValLeuValAspTrpLeuProTrp 234
Db 619 CTCGAAACATTAATGCGCATTTTC--GATGGCCCGCATCTTTCTCAGTTTCCTGCCGTGG 675
Qy 235 AsnIleThrPheGlyGlySerHisAsnIleGlyIleValLeuTyraAsnGlyIleThrTyr 254
Db 676 CATCACAGCTTTGGTGGATTAATTTGAAAAATACAGCGCTATTTTCAGTCAGCGCATGTTA 735
Qy 255 TyrLeuAspAspGlyLysProThrAlaGlnGlyPheAlaGluThrLeuArgAsnLeuSer 274
Db 736 CATATTGATATAATGCGCG-----GGCATGGATATTGACCGGATGTCACATAATCTAGT 789
Qy 275 GlnIleSerProThrAlaTyLeuThrValProLysGlyIleTrpGluGluLeuValGlyAla 294
Db 790 CAGTTAAACCGACACCTTTTTCATCGTACCCCAAGACTGTCAGTCAGTGGCCCAACAA 849
Qy 295 LeuGlnArgAspSerThrLeuArgGluArgPhePheAlaArgMetLysLeuPhePhe 314
Db 850 ATGAAACATTCGAGAGCGTTTCACCGGGCAGTGGCACGACATTCGAGATGATCTTCC 909
Qy 315 AlaAlaAlaGlyLeuSerGlnGlyIleTrpAspArgLeuAspArgValAlaGluGlnHis 334
Db 910 GCGGCCCAAAATTGGCTCGGATATCGAAGACATTTT-----AACCAT 954
Qy 335 CysGlyGluArgIleArgMetMetAlaGlyLeuGlyMetThrGluThrAlaProSerCys 354
Db 955 TGTGTG-----ATCTACGTTTGTGAGGGCTGGGGGCTGACGGAACACACCCCA---TGC 1005
Qy 355 ThrPheThrThrGlyProLeuSerMetAlaGlyTyr-----IleGlyLeu 369
Db 1006 CTGACACTTAAGGCGCG-----GGCATTAAGAACACAAACACGCTAGGCGCGC 1053
Qy 370 ProAlaProGlyCysGluValIleLeuValProValAspGlyLysLeuGlnGlyArgPhe 389
Db 1054 CTTTACCCCAATGTCATTTCTCAGATTTGATGAAGATGCGGGAGATTTTGGCAAA--- 1110
Qy 390 HisGlyProHisValMetSerGlyTyrTrpArgAlaProGluGlnAsnAlaGlnAlaPhe 409
Db 1111 ---GGGCGCAATGTGTGCTGCTATTTATGCGGATGAACCCGTAACCGCTGGCTTC 1167
Qy 410 AspGluGlnGlyTyrTyrCysSerGlyAspAlaIleLysLeuAlaAspProAlaAspPro 429
Db 1168 ---AGGAGGGATGTTTCAGAACCGGCGATCTCGGCAAGATGATC----- 1209

```

```

Qy 430 GlnLysGly---LeuMetPheAspGlyArgIleAlaGluAspPheLysLeuSerSerGly 448
Db 1210 ---AAGGTAACTTGTGCTTTCGGGCGGTTTGATTCAGTCAAAAAGCTGTATATGCC 1266
Qy 449 ValPheValSerValGlyProLeuArgThrArgAlaValLeuGluGlySerTyrVal 468
Db 1267 GAAAGGTATCCAGCGATGCAATGGA--CAAGTCTGTTGAGAAACCGATATTATC 1323
Qy 469 LeuAspValValAlaAlaProAspArgLysLeuGlyLeuLeuValPhe----- 486
Db 1324 AACCATCCATTATGATATGAAACAGCCCTCAGCTACGCTTATTTATTATTAAC 1383
Qy 487 ProArgLeuLeuAspCysArgAlaLeuSerGlyLeuGlyLysGluAlaSerAspAlaGlu 506
Db 1384 CTTGAGCTG-----TGTCCGACGATTTGGT-----GAGAAATGCACA 1422
Qy 507 ValLeuAlaSerGluProValArgAlaIleTrpPheAlaAspTrpLeuLysArgLeuArg 526
Db 1423 CTCGACACATCATCCGTTTGCAGAGCATATTTGCTGAGGTGTACACAGCATATATCAG 1482
Qy 527 GluAlaThrGlyAsnAlaSerArgIleMetTrpValGlyLeuLeuAspThrProProSer 546
Db 1483 CAGGTGAGGATAGTTCCTTTAAATGATTCGCGTGCCTGCTACTGAACCTCACTTCT 1542
Qy 547 IleAspLysGlyGluValThrAspLysGlySerIleAsnGlnArgAlaValLeuGlnTrp 566
Db 1543 CTGAGAGAAAGGGAAGTCAACCGCTATTAAAGTTCGACAGGTACTGACCGAGAAAT 1602
Qy 567 ArgSerAlaLysValAspAlaLeuTyr 575
Db 1603 TATCAGGCTGTGCGTATGCTTATAC 1629

RESULT 12
US-10-156-761-5544
? Sequence 5544, Application US/10156761
? Publication No. US20030119018A1
? GENERAL INFORMATION:
? APPLICANT: OMURA, SATOSHI
? APPLICANT: IKEDA, HARUO
? APPLICANT: ISHIKAWA, JUN
? APPLICANT: HORIKAWA, HIROSHI
? APPLICANT: SHIBA, TADAYOSHI
? APPLICANT: SAKAKI, YOSHIYUKI
? APPLICANT: HATTORI, MASAHIRA
? TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
? FILE REFERENCE: 249-262
? CURRENT APPLICATION NUMBER: US/10/156,761
? CURRENT FILING DATE: 2002-05-29
? PRIOR APPLICATION NUMBER: JP 2001-204089
? PRIOR FILING DATE: 2001-05-30
? PRIOR APPLICATION NUMBER: JP 2001-272697
? PRIOR FILING DATE: 2001-08-02
? NUMBER OF SEQ ID NOS: 15109
? SEQ ID NO 5544
? LENGTH: 1908
? TYPE: DNA
? ORGANISM: Streptomyces avermitilis
? NAME/KEY: CDS
? LOCATION: (1)..(1908)
US-10-156-761-5544

Alignment Scores:
Pred. No.: 8 31e-27 Length: 1908
Score: 335.50 Matches: 156
Percent Similarity: 39.12% Conservative: 101
Best Local Similarity: 23.74% Mismatches: 266
Query Match: 10.89% Indels: 134
DB: 15 Gaps: 26

US-09-750-986D-30 (1-589) x US-10-156-761-5544 (1-1908)
Qy 8 LeuProPheProGlyArgIleLeuGluArgLeuGluHisIleTrpAla-IleThrArgProGlu 27

```



```

QY 85 PheGlyAlaMetTyrAlaGlyIleProTyrCysProValSerProAlaTyrSerLeuLeu 104
DB 241 CTTGCGGCCGACATGTCGTATCGCCTCGGCTGCATCACGCGGCTTATGCC---ACG 297
QY 105 SerGlnPheLeuAlaValPheLeuArgHisIleValGlyLeuLeuGlnProGlyLeuValPhe 124
DB 298 TCGGCCCATCTCACCAACTCGCGCAATCGCGGGCGGATCACCCCGGGATGGTTTT 357
QY 125 AlaIleAspAlaIleProPheGlnArgAlaIleGlnThrIleLeuProAspAspValPro 144
DB 358 GCCAGGATGCGAGCGCCCTTTCGCGGGCGCTCGCGGAGTTTCGACGACGGTACACCG 417
QY 145 AlaIlePheThrArgGlyGlyLeuAlaGlyArg--ArgThrValSerPheAspSerLeu 163
DB 418 CTGGTCGGCCCTCCGCAACCTCGCGAAGATGTTCCAAATACGTTTCAATTCGAAACCTCG 477
QY 164 LeuGlnGlnProGlyGlyIleGlnAlaAspAsnAlaPheAlaIleThrGlyProAspThr 183
DB 478 CTTGAAACGGAGCGGACCGGACCGGATTGCTGCTTCATGCGGCGCGCGGATACCC 537
QY 184 IleAlaIysPheLeuPheThrSerGlySerThrLysLeuProLysAlaValProThrThr 203
DB 538 GTGCCCAATTCCTTTTACGCTCAGGCAAGACTGTTCCGCCAAGCGGTCATCCAGACG 597
QY 204 GlnArgMetLeuCyValaAsnGlnIleMetLeuLeuGlnThrPheProValPheGlyGly 223
DB 598 CAACGCAATGTTGCTCCCAATCAGGAATGATGTCGATGTTATGCTACTTTCCGAG 657
QY 224 GlnProProValLeuValAspTrpLeuProTrpAsnIleThrPheGlyGlySerHisAsn 243
DB 658 GAGCGCGCGCTGCTGTTGATGGGCGCCACGAGAACACACGCGCGGGAGAACAGTG 717
QY 244 IleGlyIleValLeuTyrAsnGlyGlyThrTyrTyrLeuAspAspGlyLysProThrAla 263
DB 718 TTCAACTGCTGCTTATATGCGGTACCTATATATCATGCGCGGACCCGACGCGCG 777
QY 264 GlnGlyPheAlaGlnThrLeuArgAsnLeuSerGlnIleSerProThrAlaTyrLeuThr 283
DB 778 GCGCAGATCGGCGACACTGACATCTCGGACCAATATCGCCACCTGGTATTTCAC 837
QY 284 ValProLysGlyTyrProGlnGlnLeuValGlyAlaLeuGlnLysArgAspSerThrLeuArgGlu 303
DB 838 GTTCCCGCTGCGCAAGAAATGCTGTCAGGCAATGCGCAAGGACGAGCGCTGTGCGC 897
QY 304 ArgPhePheAlaArgMetLysLeuPhePhePheAlaIleAlaGlyLeuSerGlnGlyIle 323
DB 898 AGCTTTTTCGCGACCTGAAGATCTGATGTATGCTCGGTCGCGCATGGCGACACACA 957
QY 324 TrpAspArgLeuAspArgValAlaGlnGlnHisCysGlyGlnLysArgIleArgMetMetAla 343
DB 958 TGGGATGCGGTGACGAGGACTTTCATGCAACGCGTGGCCATGGCTTCATCGATGCC 1017
QY 344 GlyLeuGlyMetThrGlnThrAlaProSerCysThrPheThrThrGlyProLeuSerMet 363
DB 1018 GGCTCTGCTGACGAGGACGCGACCTTCTGCTGTTCGACCGGAACCGGACGACAG 1077
QY 364 AlaGlyTyrIleGlyLeuProAlaProGlyCysGlnValLysLeuValProValAspGly 383
DB 1078 CCGGCAATTCGCTATCGGCGGCGGAGGCTGACGATGAATCGTGCCCTTCGATGCGC 1137
QY 384 LysLeuGlnGlyArgPheHisGlyProHisValMetSerGlyTyrTrpArgAlaProGln 403
DB 1138 CCGTACGAAATTCAGCGCTGAAGCGGACATACGCGCGTTATCGCTAATGGGAA 1197
QY 404 GlnAsnAlaGlnAlaPheAspGlnGlyTyrTyrCysSerGlyAspAlaIleLysLeu 423
DB 1198 TTGACGCGCGGACGCTTCGACGAGGAGGCTTCATGAATCGGCGATTCGCAAAATTC 1257
QY 424 AlaAspProAlaAspProGlnLysGlyLeuMetPheAspArgLysIleAlaGlnAspPhe 443
DB 1258 GCGGTGCGGACGATCGCGCGCGGCTTTTATTCATGACGCGCATGGCGGAAACTTC 1317

```

```

QY 444 LysLeuSerSerGlyValPheValSerValGlyProLeuArgThrArgAlaValLeuGln 463
DB 1318 AAGCTGCAGACCGGACCTGGGTGGCGGTGGCGCTTTCGCGGCAATTAGTCAACATG 1377
QY 464 GlyIlySerTyrValLeuAspValValAlaIleAspAspArgGlyCysLeuGlyLeu 483
DB 1378 TTCGCGGGCTGATCCCTGATGCGGTATACGCGGCGAAGAACCGGCGGATGGGGCT 1437
QY 484 LeuValPheProArgLeuLeuAspCysArgAlaLeuSerGlyLeuGlyLysGlnAlaSer 503
DB 1438 CTGCTGTCGTCATTCAATCCGCGGCTACGTGAGCTTGTGGCGTTGGCAATCTCTCC 1497
QY 504 AspAlaGlnValLeuAlaSerGlnProValArgAlaTrpPheAlaAspTrpLeuLysArg 523
DB 1498 GATCGGAGATCATCCGCGACCCGAGGTGTCAGGGCGAGATTGCGGAAGCTTCGGCG 1557
QY 524 LeuAsnArgGlnAlaThrGlyAsnAlaSerArgIleMetTrpValGlyLeuLeuAspThr 543
DB 1558 CATCAAAACAGGAGCGGCTTCGCAAGCGCGCTCATGCGCATTCGTGATGAAGAC 1617
QY 544 ProProSerIleAspLysGlyValThrAspLysGlySer 557
DB 1618 GC-GCTGCGCTTCGAGAGCGAGGATTACGCAAGAGGATCG 1658

```

```

RESULT 10
US-10-369-493-40734
; Sequence 40734, Application US/10369493
; Publication No. US2003023675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OR INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 40734
; LENGTH: 1866
; TYPE: DNA
; ORGANISM: Caulobacter crescentus
US-10-369-493-40734

Alignment Scores:
Pred. No.: 8,89e-98 Length: 1866
Score: 972.00 Matches: 219
Percent Similarity: 53.32% Conservative: 86
Best Local Similarity: 38.29% Mismatches: 249
Query Match: 31.54% Indels: 18
DB: Gaps: 6

US-09-750-986d-30 (1-589) x US-10-369-493-40734 (1-1866)
QY 11 ProGlyArgIleLeuGlnArgLeuGlnLysTrpAlaLysThrArgProGlnGlnThrCys 30
DB 148 CCGCGCTCGATCCGCCACCTGATGCGCGGCGGCGGCGGCGGACATCCGCGGCTTAC 207
QY 31 ValAlaAlaArgAlaIleAsn---GlyGlnTrpArgArgIleSerTyrAlaGlnMetPhe 49
DB 208 CTAGACGCGGCAACCGAACGACGAGTCCCTGGCGGCGGCTCACTATGGCGAGGCGCAT 267
QY 50 HisAsnValArgAlaIleAlaGlnSerLeuLeuProTyrGlyLeuSerAlaGlnArgPro 69
DB 268 CCGCGGCTGAGGCGGCTCGCGGCGGCTGATGATCAGAGCTTCAGGCCCGACGACAGC 327
QY 70 LeuLeuIleValSerGlyAsnAspLeuGlnHisLeuGlnLeuAlaPheGlyAlaMetTyr 89
DB 328 GTGATGATCTCTGCGGCAATTGATCGAGCAGCGCCTGATGACGCTGGCGGCTTATACG 387

```

```

QY 105 SerGlnAspLeuAlaIysLeuArgHisIleValGlyLeuLeuGlnProGlyLeuValPhe 124
DB 298 TCGCCGATCTCCAAACTCGCGACATCGCGGGAGATCACTCGGGAGATGGTTT 357
QY 125 AlaAlaAspAlaIaProPheGlnArgAlaIleGluThrIleLeuProAspAspValPro 144
DB 358 GCGGAGATCGACGCCCTTTCGCCGGCCCTCGCGCAAGTTTTCGACGACGGTACACCG 417
QY 145 AlaIlePheThrArgGlyLeuAlaGlyArg--ArgThrValSerPheAspSerLeu 163
DB 418 CTGGTCGGCTCCGCAACTCGCGGAAGATGTTCCAAATGCTTCATTTCGAAACCTCG 477
QY 164 LeuGluGlnProGlyIleGluAlaAspAspAlaPheAlaIaThrGlyProAspThr 183
DB 478 CTGGAACGAGCGACCGAAGCGTTGATGTCCTTCATCGCGTGGCGCGGATACC 537
QY 184 IleAlaIysPheLeuPheThrSerGlySerThrIleLeuProIysAlaValProThrThr 203
DB 538 GTGCGCAATCTCTTTTCACTGACGACGACGATGTTCCGCAAGCGTCAATCCAGACG 597
QY 204 GlnArgMetLeuCysAlaAsnGlnMetLeuGlnThrPheProValPheGlyGlu 223
DB 598 CAACGATGTTGCTCAATCAGAAATGATTGCGCATTTGACTCTTCGCGAG 657
QY 224 GluProProValLeuValAspThrLeuProThrAsnIleThrPheGlyIysSerHisAsn 243
DB 658 GACCCCGCGGTCTGTTGATGGCGCCCATGAAACACCGCGCGGGAACAGGTG 717
QY 244 IleGlyIleValLeuIleuTyraGlnGlyIleThrTyreIleuAspAspGlyIleProThrAla 263
DB 718 TTCACTCTCTCTCTATATGCGGACCTATATATGATCGCGCAAGCCAGCCG 777
QY 264 GlnGlyPheAlaGluThrLeuArgAsnLeuSerGluIleSerProThrAlaTyreLeuThr 283
DB 778 GCGCAGATCGGCACAGACTGACCAATCTCGGACATATCGCCACCTGGTATTTCAAC 837
QY 284 ValProIleGlyIleThrGluLeuValGlyAlaLeuGluArgAspSerThrLeuArgGlu 303
DB 838 GTTCCCGCTGCGCACAGAAATGCTGCTCAGCAAGATGCGCAAGCGCGCTGTCGCC 897
QY 304 ArgPhePheAlaArgMetIleLeuPhePhePheAlaIaIaGlyLeuSerGlnGlyIle 323
DB 898 AGCTTTTTCGCGACCTGAAGATGCTGATGATGCGGTCCGCGCATGGCGACACACA 957
QY 324 TyrAspArgLeuAspArgValAlaGluGlnHisCysGlyIleuArgIleArgMetMetAla 343
DB 958 TGGGATGCGCTGACGAGACTTTCATGCAACGCTCGGCATGCGCTTCATGCGTGGC 1017
QY 344 GlyLeuGlyMetThrGluThrAlaProSerCysThrPheThrArgIleProLeuSerMet 363
DB 1018 GGCCTGCGCTCGACGAGACGCGACCTTCTCGCTGTTTCGACCGAAGCGCGACAG 1077
QY 364 AlaGlyTyrIleGlyLeuProAlaProGlyCysGluValIleuValProValAspGly 383
DB 1078 CCGGGAATTCGGTATCCCGGCGCGAGGGGTGACATGAATCGTGGCTTCGATGGC 1137
QY 384 IysLeuGluGlyArgPheHisGlyProHisValMetSerGlyTyrTrpArgAlaProGlu 403
DB 1138 CGGTACCAATTAACGCGTGAAGGCCGACATCAACGCGGTACTGCGCTAATGCGGAA 1197
QY 404 GlnAspAlaIleAlaPheAspGluGluGlyTyrTrpCysSerGlyAspAlaIleValLeu 423
DB 1198 TTGACGCGGACGCTTCGACGAGAGGGTCTCATGAATCGCGCATACCGTAAATTT 1257
QY 424 AlaAspProAlaAspProGluIleGlyLeuMetPheAspGlyArgIleAlaGluAspPhe 443
DB 1258 GCGGTGCGGACGATCGCGCGCGCTTTTATTTGATGCGCGCATGGCGGAAACTTC 1317
QY 444 IysLeuSerSerGlyValIleValSerValGlyProLeuArgThrArgAlaValLeuGlu 463
DB 1318 AACGTGACAGCGGACACTGGGTGGCGGTGGCGCTTTCGCGCAATTAGTCMACATG 1377

```

```

QY 464 GlyIleSerTyraIleuAspValValAlaIaProAspArgGluCysLeuGlyLeu 483
DB 1378 TTGCGCGGCTGATCCGTATGTCCTCATACGCGGAGAAACCGCGCAACTGGCGCT 1437
QY 484 LeuValPheProArgLeuLeuAspCysArgAlaLeuSerGlyLeuGlyIleValAsp 503
DB 1438 CTGCTGCTGATTCATACCGCGCGCTACGTACGACTGTTTCGCGTTTCGCAACATCTCC 1497
QY 504 AspAlaGluValLeuAlaSerGluProValIleArgAlaIleThrPheAlaAspThrLeu 523
DB 1498 GATCGGAGATCTATCCGCGCACCCGAGCTCAGGCGCCAGATTGTCGCAAGCTTTCGCG 1557
QY 524 LeuAspArgGluAlaThrGlyAsnAlaSerArgIleMetTrpValGlyLeuLeuAspThr 543
DB 1558 CATCAAAAACAGCGAGCGGCTTCGCAAGCGCGTATCGCATTCCTCGATGAGAAC 1617
QY 544 ProPheSerIleAspIysGlyIleValIleAspIysGlySerIleAsnGlnArgAlaVal 563
DB 1618 GCGCTGCGCTTCGAGAAAGCGAGTTTACGACAAAGGATCGATCAACGACGCGCGTG 1677
QY 564 LeuGlnTrpArgSerAlaIysValaIaAspAlaLeuTyr 575
DB 1678 CTTCGATCGCAAGAGCTCGTAGATCACTTAC 1713

RESULT 9
US-10-369-493-38007
; Sequence 38007, Application US/10369493
; Publication No. US2003023675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xiandeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 38007
; LENGTH: 1660
; TYPE: DNA
; ORGANISM: Agrobacterium tumefaciens
US-10-369-493-38007

Alignment Scores:
Score: 3,64e-126 Length: 1660
Percent Similarity: 61.19% Matches: 255
Best Local Similarity: 46.03% Mismatches: 84
Query Match: 39.78% Indels: 213
DB: 15 Gaps: 2

US-09-750-986D-30 (1-589) x US-10-369-493-38007 (1-1660)
QY 5 GluAlaLeuLeuProPheProGlyArgIleLeuGluArgLeuGluIleThrAlaIleThr 24
DB 1 GATCCGCTGGCCCTCATCGCAAAATTAATGAAGACGCTCTCTTCACCTGATCGTTCG 60
QY 25 ArgProGluGlnThrCysValaIaIaIaArgAlaIaIaAsnGlyCyluThrArgArgIleSer 44
DB 61 GCACCGAGGCAACATGAGTACCGGACCGTCAAGGGTCCGAGCCATGCGCGCGGTCCAGC 120
QY 45 TyraIleuMetPheHisAsnValaIaIaIaGlnIleSerLeuLeuProTyraGlyLeu 64
DB 121 TACCGGAGGCTTCGACAAATAGCGCGATGCGCATGCTCTTCGATCATGACCTT 180
QY 65 SerAlaGluArgProLeuLeuIleValSerGlyAsnAspLeuGluHisIleGluIleVal 84
DB 181 TCGGTGAGAGGCGCGCTGCTGCTTCTGAAATTCATCGAGCATGCGTTGATGGTG 240

```


QY 145 AlaIlePheThrArgGlyLeuAlaGlyArg---ArgThrValSerPheAspSerLeu 163
 Db 418 GTGGTGGCTCGTCAACCTGGGAGAGATGTTCCAAATGCTTTCATTTGAAACCTTG 477
 QY 164 LeuGlnProGlyGlyIleGluAlaAspAsnAlaPheAlaIleThrGlyProAspThr 183
 Db 478 CTTGAAACGAGCGGACCGGAGGCTTGGTCTTCATGCGGCGGCGGATACC 537
 QY 184 IleAlaPhePheLeuPheThrSerGlySerThrIleLeuProIleValIleProThrThr 203
 Db 538 GTGCGCAAAATTCCTTTTCATGTCAGGACGACATGCTGCGCCAAAGGGCTCATCCAGACG 597
 QY 204 GlnArgMetLeuGlyAlaIleGlnIleMetLeuGlnIleThrPheProValPheGlyLeu 223
 Db 598 CAACGAGATGTTGCTTCAATCAGAAATGATGCGCATTTGAGCTTTCGCGAG 657
 QY 224 GluProProValLeuValaAspTrpLeuProTrpAsnIleThrPheGlyGlySerHisAsn 243
 Db 658 GAGCCCGGCTCGTCTGATTTGGCGCCATGGAACACACGCGCGGGAACAGGTG 717
 QY 244 IleGlyIleValIleuTyraGlyGlyIleThrTyrrTyrrLeuAspArgIlyLysProThrAla 263
 Db 718 TTCACCTCGCTCTTATATGGGGGTACCTATATATGATGCGCGCAAGCCAGCGCG 777
 QY 264 GlnGlyPheAlaGluThrLeuArgAsnLeuSerGluIleSerProThrAlaIleLeuThr 283
 Db 778 GCGCAGATCGCGACACATGACATCTCGGACATATCGCCCATCTGATTTTCAC 837
 QY 284 ValProIleGlyIleTrpGluIleValaGlyAlaIleGluIleArgAspSerThrLeuArgGlu 303
 Db 838 GTTCCCGCTGGCACGAAATGCTGCTCCAGGCAATGCGCAAGACAGCGGCTGTGCGC 897
 QY 304 ArgPhePheAlaArgMetIleLysPhePheAlaAlaIleGlyLeuSerGlyIle 323
 Db 898 AGCTTTTTCGCGACCTGAAGATGCTGATGATGCGGCTCGGCAATGGCGGACAGACACA 957
 QY 324 TrpAspArgLeuAspArgValaIleGluIleHisCysGlyLeuArgIleArgMetAla 343
 Db 958 TGGGATGCGCTGCGGAGCTTTCATGCAACGCTGCGCATGCGCTTCATGCGGTGCC 1017
 QY 344 GlyLeuGlyMetThrGluThrAlaProSerCysThrPheThrThrGlyProLeuSerMet 363
 Db 1018 GGCCTCGGCTCGACGAGACGACCTTCTCGCTGTTCTGCAACCGCAACCGGACGACAG 1077
 QY 364 AlaGlyTyrrIleGlyLeuProAlaProGlyCysGlyValIleLysLeuValProValAspGly 383
 Db 1078 CCGGCAATATCGGTATCCCGGCGAGGCGTGACATGAACCTGCGCTTCGATGGC 1137
 QY 384 LysLeuGlnGlyArgPheHisGlyProHisValMetSerGlyTyrrTrpArgAlaProGlu 403
 Db 1138 CGGTACGAAATTCAGCGTGAAGGCGCCGACATCAGCCCGGTATCTGGCTAATGGGAA 1197
 QY 404 GlnAsnAlaGlnAlaPheAspGlnGlyTyrrTyrrCysSerGlyAspAlaIleLysLeu 423
 Db 1198 TTACGCGGCGACGCTTTCGACGAGAGGCTTCTATGCAATCGGCATACCGTCAATTT 1257
 QY 424 AlaAspProAlaAspProGlnIleGlyLeuMetPheAspGlyArgIleAlaGluAspPhe 443
 Db 1258 GCGCTCGCGAGCATCGCGCGCGCTTTATTTGATGCGCGCATGGCGGAAACTTC 1317
 QY 444 LysLeuSerSerGlyValPheValSerValGlyProLeuArgThrArgAlaValLeuGlu 463
 Db 1318 AACCTGAGACCGGACACTGCGTGGCGCTGCGGCTTTCGCGGCAATTAATGCAACATG 1377
 QY 464 GlyGlySerTyrrValLeuAspValaIleValaIleAlaProAspArgGlyCysLeuGlyLeu 483
 Db 1378 TTGGCGGCGCTGATCGTGTATGCGCTCATTAACGCGGCAAAACCGGCGCACTGGCGCT 1437
 QY 484 LeuValPheProArgLeuAspCysArgAlaLeuSerGlyLeuGlyLysGluAlaSer 503
 Db 1438 CTGCTGCTGCTCATTAACGCGGCGCTACGTGAGCTTCTGCGGCTTCGCAACATCTCTCC 1497

QY 504 AspAlaGlyValIleLeuAlaSerGluProValArgAlaTrpPheAlaAspTrpLeuLysArg 523
 Db 1498 GATCGGAGATCATTCGCGCACCGAGTGTACAGGCGCCCAATTCCTCCGAAAGCTTTGGCG 1557
 QY 524 LeuAsnArgGlnAlaThrGlyAsnAlaSerArgIleMetTrpValGlyLeuLeuAspThr 543
 Db 1558 CATCAAAACAGCGGCGGCTTCGCGACCGCGTCATGCGGCTTCGATGATGAAGAC 1617
 QY 544 ProProSerIleAspIleGlyValIleThrAspIleGlySerIleAsnGlnAlaVal 563
 Db 1618 GCGCTCGCTTCGAGAAAGCGGCTTACGAGACAAAGGATCGATCAACAGCGCGCGTG 1677
 QY 564 LeuGlnTrpArgSerAlaLysValaAspAlaLeuTyr 575
 Db 1678 CTTCTGATTCGCAAGAGCTCTTAAGTCACTTTAC 1713
 RESULT 7
 US-10-369-493-38427
 ; Sequence 38427, Application US/10369493
 ; Publication No. US2003023675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; PRIOR FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 38427
 ; LENGTH: 1728
 ; TYPE: DNA
 ; ORGANISM: Agrobacterium tumefaciens
 ; US-10-369-493-38427
 Alignment Scores:
 Pred. No.: 1,01e-133 Length: 1728
 Score: 1294.00 Matches: 266
 Percent Similarity: 61.19% Conservative: 84
 Best Local Similarity: 46.50% Mismatches: 220
 Query Match: 41.99% Gaps: 2
 DB: 15
 US-09-750-986d-30 (1-589) x US-10-369-493-38427 (1-1728)
 QY 5 GlnAlaLeuLeuProPheProGlyArgIleLeuGlnArgLeuGluHisIleTrpAlaLysThr 24
 Db 1 GATCGCTTGGCCCTTCATCGGCAAAATTAATGAATGAGACGCTCTTCATCGTGTCTCG 60
 QY 25 ArgProGlnIleThrCysValaIleAlaArgAlaAlaAsnGlyLysIleTrpArgArgIleSer 44
 Db 61 GCACCGAGGCGACATGAGATGCGCGACGTCAGGCTCGGACCATGCGCCGCGGTCAAC 120
 QY 45 TyrAlaGluMetPheHisAsnValArgAlaIleAlaGlnSerLeuLeuProTyrrGlyLeu 64
 Db 121 TACCGGAGGCTTCGACAAAGATACGCGGTATGGCGCAAGTCTTCTCGATCATGACCTT 180
 QY 65 SerAlaGluArgProLeuLeuIleValSerGlyAsnAspLeuGluHisIleGlnIleAla 84
 Db 181 TCGGTGAGAGGCGCGTGTGCTTCTGAAAATTCATCAGCATGCGATGCGTGTGATGCTG 240
 QY 85 PheGlyAlaMetTyrrAlaGlyIleProTyrrCysProValSerProAlaTyrrSerLeuLeu 104
 Db 241 CTTGCGGCGAGCATGTGCTATGCGCTGCGCTGCGCATACGCGGCTATGCTC---AGC 297
 QY 105 SerGlnAspLeuAlaLysLeuArgHisIleValGlyLeuLeuGlnProGlyLeuValPhe 124
 Db 298 TCGGCGGATCTCAACAACTCGCGACATCCGCGGCGAGATCAACCCCGGAGATGTTT 357

QY 161 AspSerLeuLeuGluGlnProGlyGlyIleGluValAspAsnAlaPheAlaIleThrGly 180
 Db 481 GCCGAACTGCTGGGAGACGCCGGAACATCCGGAATCCCGCAAGCAGCAAGCATGCGC 540
 QY 181 ProAspThrIleAlaIysPheLeuPheThrSerGlySerThrIlybLeuProGlyValAla 200
 Db 541 CACGACCAATATCGGTCTGACGCGTGTCAACGCGGCTGATGTATATCGACGCGCAG 780
 QY 201 ProThrThrGlnArgMetLeuCyAlaAsnGlnMetLeuGlnThrPheProVal 220
 Db 601 ATCAATATCGGAGGATATGCGCCATATGATATATCCGCGAGGAGATGCGCTTC 660
 QY 221 PheGlyGluGlnProValIleValAspTrpLeuProTrpAsnIleThrPheGly 240
 Db 661 CTGAAGACGACGCGCGGTGATGTCGATGCGTGCAGGATATCACCTTCGCGGCGC 720
 QY 241 SerHisAsnIleGlyIleValIleuTyraGlyGlyThrTyrrTyraAspAspGlyLys 260
 Db 721 AACCAAAATATCGGTCTGACGCGTGTCAACGCGGCTGATGTATATCGACGCGCAG 780
 QY 261 ProThrAlaGlnGlyPheAlaGluThrLeuArgAsnLeuSerGlnIleSerProThrAla 280
 Db 781 CCGACGCGCGCGGATGCTTCCACCATCCGCAATCGCGCGAGATCGCGCGCGCGTGC 840
 QY 281 TyrLeuThrValProLysGlyTrpGluGluValGlyAlaLeuGluArgAspSerThr 300
 Db 841 TATTTCAAGCTTCGAAAGGCTACGATCGCTGTCGCGGTGCGCGCGCAAGCACACAG 900
 QY 301 LeuArgGluArgPhePheAlaArgMetLysLeuPhePheAlaAlaIleGlyLeuSer 320
 Db 901 TTGCGCAAAATGTTCTTCACGCGGCTGCATGCGATGTTCTTCGCGCGCGCGCTCGCG 960
 QY 321 GlnGlyIleTrpAspArgLeuAspArgValAlaGluIleHisCysGlyGlyArgIleArg 340
 Db 961 GCCCATGTCGGAACGCGCTTGACGAGCTGCGGTCCGGAAGACGCGCGCGGTGCGC 1020
 QY 341 MetMetAlaIleuGlyMetThrGluThrAlaProSerCysThrPheThrThrGlyPro 360
 Db 1021 ATGCTCACCGGCTTGCGCGCACCGACCGCGCTTTCATATGTCGATGACCGCGCAG 1080
 QY 361 LeuSerMetAlaGlyTyrrIleGlyLeuProAlaProGlyCysGlyValLysLeuValPro 380
 Db 1081 ACCAGTCTGCGGCGCATGTCGCGCTGCGGTGCGCGCAACGAGCGCAAGCTGTGCGC 1140
 QY 381 ValAspGlyLysLeuGluArgPheHisGlyProHisIleValMetSerGlyTyrrTrpArg 400
 Db 1141 AACAAACGCGAAGCTCGAAGTCCGCGCAAGGCGCGAATACACCCCGGCTATTTGCGC 1200
 QY 401 AlaProGluGlnAsnAlaGlnAlaPheAspGluGluGlyTyrrTyrrCysSerGlyAspAla 420
 Db 1201 GCGCGCGAGCTGACCGCATTAAGGCTTTCAGACGAGGAGGCTTTCATCAAGCTCAACGATGCG 1260
 QY 421 IleLysLeuAlaAspProAlaAspProGlnLysGlyLeuMetPheAspGlyArgIleAla 440
 Db 1261 CTGAAGCGCGGTATGCAAGACCTTTCGCGCGCTTCATTCATTCGACGCGCGGATCTCG 1320
 QY 441 GluAspPheLysLeuSerSerGlyValIlePheValSerValGlyProLeuAlaGlnThrArgAla 460
 Db 1321 GAAGACTTCAAACTGCGGTGCGGACCTGCGGTCTGCGGTCTGCGGTCAACGCGCAAGTTC 1380
 QY 461 ValLeuGluGlyGlySerTyrrValIleuAspValValAlaAlaProAspArgGlyCys 480
 Db 1381 ATTGCGCGCTGCGCTCTCTGATGCGGATGATGATGCGCGGCTGACGCGGATATAC 1440
 QY 481 LeuGlyLeuLeuValIlePheProArgLeuLeuAsp-----CysArgAlaLeuSerGly 497
 Db 1441 GTACACGCGGTGCG-----ATCCTGATCCGACGCGCTGCAAGCTGATCAATGCG 1491
 QY 498 LeuGlyLysGluAlaSerAspAlaGluValLeuAlaSerGluProValAlaArgAlaTrpPhe 517
 Db 1492 ACCCTGCGCTGGAAGACTCTGCGCGGATGCGCGCATCTGCATCCGCGAGGCGCTTC 1551

QY 518 AlaAspTrpLeuLysArgLeuAsnArgGluAlaIleThrGlyAsnAlaSerArgIleMetTrp 537
 Db 1552 CGCGAGCGCTTCCGCCACGCTGTCGACGAGCGCACCGGCTCTGTCACACCGCGCACCGC 1611
 QY 538 ValGlyLeuLeuAspThrProProSerIleAspLysGlyGluValAlaThrAspLysGlySer 557
 Db 1612 GCGGTGCTGCTCGGCGAACCCTGTCATGACAGAGGAGAGATCAACGCAAGGAGCTCG 1671
 QY 558 IleAsnGlnArgAlaValLeuGlnTrpArgSerAlaLysValAspAlaLeuTyrr 575
 Db 1672 GTCAACAGCGGCGCGCTGCGAATATCGCGCTCGTTCATCCGAGATCTTTAC 1725
 RESULT 6
 US-10-369-493-35384
 ; Sequence 35384, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 ; FILE REFERENCE: 38-10 (52052) B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; PRIORITY FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 35384
 ; LENGTH: 1728
 ; TYPE: DNA
 ; ORGANISM: Agrobacterium tumefaciens
 US-10-369-493-35384
 Alignment Scores:
 Pred. No.: 1.01e-133 Length: 1728
 Score: 1294.00 Matches: 266
 Percent Similarity: 61.19% Conservative: 84
 Best Local Similarity: 46.50% Mismatches: 220
 Query Match: 41.99% Indels: 2
 DB: 15 Gaps: 2
 US-09-750-986d-30 (1-589) x US-10-369-493-35384 (1-1728)
 QY 5 GlnAlaLeuLeuProPheProGlyArgIleLeuGluArgLeuGluHisTrpAlaLysThr 24
 Db 1 GATCCGCTTGCGCCCTATCCGCAAAATTAAGATGAGGTCTCTTCACTGCTGTGCTTCG 60
 QY 25 ArgProGluGlnThrCysValAlaAlaArgAlaAlaAsnGlyGluTrpArgArgIleSer 44
 Db 61 GCACCGGAGGCAATGATGATGATGCGGACCGTCAAGGCTGCGAAGCATGCGCGCGGTACG 120
 QY 45 TyrAlaGluMetPheHisAsnValArgAlaAlaAlaGlnSerLeuLeuProTyrrGlyLeu 64
 Db 121 TACGCGAGGCTCTCGACAAATACGCGCGATGCGGCAAGTTCCTTCATCATGACCTT 180
 QY 65 SerAlaGluArgProLeuLeuIleValSerGlyAsnAspLeuGluHisLeuGlnLeuAla 84
 Db 181 TCGGTCCAGAGCGCGCTGCTGCTGCTTTCGAAATTCCTTCAGACGTGCGTGTGATGCTG 240
 QY 85 PheGlyAlaMetTyrrAlaGlyIleProTyrrCysProValSerProAlaTyrrSerLeu 104
 Db 241 CTTCGCGCGCAGCATGTGCTATGCTGCTGCGCTGCGCTGCAACGCGGCGCTATAGCC--ACG 297
 QY 105 SerGlnAspLeuAlaLysLeuAlaArgHisIleValGlyLeuLeuGlnProGlyLeuValPhe 124
 Db 298 TCGGCGCATCTTCAACAACTCGCGCATCGCGGAGAGATCAACCCGCGGAGATGTTT 357
 QY 125 AlaAlaAspAlaAlaProPheGlnArgAlaIleGluThrIleLeuProAspArgValPro 144
 Db 358 GCCGAGATGCAAGCCCTTTCGCGGCGCTGCGGAGGATTTTCAGACGAGTACACCG 417

QY 201 ProThrThrGlnArgMetLeuCyAlaAsnGlnGlnMetLeuGlnInThrPheProVal 220
DB 21888 CCGACTACTAGGAGGAGTCTGTGGCCAAATCAGAGATGCTTCTGCAAACTTCTCCGGTT 21947
QY 221 PheGlyGlnGlnProProValLeuValAspTrpLeuProTrpAsnHisThrPheGly 240
DB 21948 TTGGTGAAGAGCGCGGCTGTGTGAGTGTGCTGTGCGTGGAAACACACTTCCGGCGC 22007
QY 241 SerHisAsnIleGlyIleValLeuTyrAsnGlyGlyThrTyrTyrLeuAspAspGlyLys 260
DB 22008 AGCCACAACTCGCATCTGTGTGTACAGAGCGCGCATCTACTTGTGACGACGCTAAA 22067
QY 261 ProThrAlaGlnGlyPheAlaGlnThrLeuAspAsnSerGluIleSerProThrAla 280
DB 22068 CCAACGCGCCAGAGGCTTGGCCAGACCTTGGCAACTTGAAGCAAACTCTCCACACTGCG 22127
QY 281 TyrLeuThrValProLysGlyTrpGlnGlnValGlyAlaLeuGlnLysAspSerThr 300
DB 22128 TACCTACTGTGGCGAAAGCTGGAGAGATTTAGTGGTGGCTTGAAGCGAGACAGTACC 22187
QY 301 LeuArgGlnArgPhePheAlaArgMetLysLeuPhePhePheAlaAlaAlaGlyLeuSer 320
DB 22188 CTGCGCGAAGCTTCTTCTGCGCATGACCTGTTCTTCTGCGCGCGCTGGGTTGCTCG 22247
QY 321 GlnGlyTyrLeuAspArgLysAspArgValAlaGlnGlnHisCysGlyGlyLysArg 340
DB 22248 CAAAGGAGTCTGGAGCTTGGACCGGGTGGCTGAACGACCTGTGGTGAAGCGCATTCG 22307
QY 341 MetMetAlaGlyLeuGlyMetThrGlnThrAlaProSerCysThrPheThrThrGlyPro 360
DB 22308 ATGATGGCGGGTCTGGCGCATGAGAGATGCTCTTCTGCACTTTTACACCGGACCG 22367
QY 361 LeuSerMetAlaGlyTyrIleGlyLeuProAlaProGlyCysGlyLysValLeuValPro 380
DB 22368 CTGTCGATGCTGTGATCATTTGGGCTGCCAGCGCTGGCGAGGCTCGTTCCG 22427
QY 381 ValAspGlyLysLeuGlnGlyLysArgPheHisGlyProHisValMetSerGlyTyrTrpArg 400
DB 22428 GTGAGTGGAGAAATTTGAGAGGCGTTTTCATGTGTCGCGCATGACAGCGGCTACTGGCGT 22487
QY 401 AlaProGlnGlnAsnAlaGlnAlaPheAspGlnGlnGlyTyrTyrCysSerGlyAspAla 420
DB 22488 GCTCTGAACAAATGCGCAAGCTTGCAGAGAGGCTTATTCCTCCGGGATGCC 22547
QY 421 IleLysLeuAlaAspProAlaAspProGlnLysGlyLeuMetPheAspGlyLysAla 440
DB 22548 ATCAAAATTTGCGAGATCTGCGCGATCTCAGAAAGCTGTGATGTTGACGCGTCGATTCGCT 22607
QY 441 GlnAspPheLysLeuSerSerGlyValPheValSerValGlyProLeuAlaGlnThrArgAla 460
DB 22608 GAAAGACTTCAAGCTGTCTCAGAGGAGTATTTGTCAAGCTGTGGCCATTTGCGACGCGGCGC 22667
QY 461 ValLeuGlnGlyGlySerTyrValLeuAspValValAlaAlaProAspArgGluCys 480
DB 22668 GTTCTGGAAGCGGCTTTCAGTCTGTGAGAGTGTGCTGTGCTGATTCGTAATGC 22727
QY 481 LeuGlyLysLeuValPheProArgLeuLeuAspCysArgAlaLeuSerGlyLysGlyLys 500
DB 22728 CTGGATGCTGTGCTGTGCTGCGGCTTCTGCACTGCGGCTGTGTGTGGGGCTAGGAAA 22787
QY 501 GlnAlaSerAspAlaGlnValLeuAlaSerGlnProAlaArgAlaTrpPheAlaAspTrp 520
DB 22788 GAGCGCTGGAGCGCGAGTGTGCGAGTGAAGCGGCTGGGCTGTGCTGTAATGC 22847
QY 521 LeuLysArgLeuAsnArgGlnAlaThrGlyAsnAlaSerArgIleMetTrpValGlyLeu 540
DB 22848 CTCAAAGACTCATTCAGAGAGCACTGGCAATGCCAGTCGCAATCATGTGGTGAAGCTC 22907
QY 541 LeuAspThrProProSerIleAspLysGlyGlyValThrAspLysGlySerIleAsnGln 560
DB 22908 CTGAGATGCGCGCTGTGATTTGATTAAGGCGAGGCTCATCAAGAGGCTGATTCACAG 22967

QY 561 ArgAlaValLeuGlnInThrPargSerAlaLysValAspAlaLeuTyrArgGlyLysArgGln 580
DB 22968 CGCGCTGTTTGGAAATGGCGGTGGCGAAAGTTGATGCCCTGTATGTGTGAAGATCAA 23027
QY 581 SerMetLeuArgAspGlnAlaThrLeu 589
DB 23028 TCCATGCTGCTGAGCAGAGGCCACACTG 23054
RESULT 5
US-10-369-493-44177
; Sequence 44177, Application US/10369493
; Publication No. US2003023675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ. ID NOS: 47374
; SEQ. ID NO 44177
; LENGTH: 1725
; TYPE: DNA
; ORGANISM: Rhodospseudomonas palustris
US-10-369-493-44177

Alignment Scores:

Pred. No.: 6.8e-152 Length: 1725
Score: 1457.00 Matches: 292
Percent Similarity: 65.40% Conservative: 86
Best Local Similarity: 50.52% Mismatches: 194
Query Match: 47.27% Indels: 6
DB: 15 Gaps: 2

US-09-750-986D-30 (1-589) x US-10-369-493-44177 (1-1725)

QY 1 MetArgSerLeuGlnAlaLeuLeuProPheProGlyArgIleLeuGlnArgLeuGlnHis 20
DB 1 GTCGCTCCAGCACACACACGCTCACCGACTATCCGTCGGATTTACGACCGGCTGCATCAC 60
QY 21 TrpAlaLysThrArgProGlnGlnInThrCysValAlaAlaArgAlaAlaAsnGlyLysTrp 40
DB 61 TTCGCGAGACGGCGCGCGACCGGCTGTTCATGTGCGCGAGCGGAAAGCTGAGGCGGCTGG 120
QY 41 ArgArgIleSerTyrArgIleGlnMetPheHisAsnValArgAlaIleAlaGlnSerLeuLeu 60
DB 121 CGCAGATCAGCTATCCCGAGATGCTGCCGCGCGCCAGACCAATCCCTCGGCTGATC 180
QY 61 ProTyrGlyLeuSerAlaGlnArgProLeuLeuIleValSerGlyAsnAspLeuGlnHis 80
DB 181 GCCGCGGAGCTGCGCGCGCAAGCGCGGTATATCTGTCGGTAAATTCATGCATCACAT 240
QY 81 LeuGlnLeuAlaPheGlyAlaMetTyrAlaGlyIleProTyrCysArgProValSerProAla 100
DB 241 GCGATGTGATGTTCCGCGCGCTGTATGCGGCGGTGCGCGATGTCGCGGTGTCGCGCGC 300
QY 101 TyrSerLeuLeuSerGlnAspLeuAlaLysValArgHisIleValGlyLeuGlnInPro 120
DB 301 TATTCGTGTGTTCAGAGATTACGCGAAGCTGCGCATATCTCGGCTGCTGACTCGG 360
QY 121 GlyLeuValPheAlaAlaAspAlaAlaProPheGlnArgAlaIleGlnThrIleLeuPro 140
DB 361 GGGCTATCTTCGCGCATATGACATCATCTGCTTGGCGCGCGCATCTCCGCAACCGGCGG 420
QY 141 AspAspValProAlaIlePheThrArgGlyGlyLeuAlaGlyLysArgThrValSerPhe 160
DB 421 GAGAGTGTGAGCTCGCTGCCAGCGCGCGGAGGTGAAGGGCGCAAGGTGACGTGCTT 480

```
QY 261 ProThraIaGInGlyPheAlaGluThrLeuArgAsnLeuSerGluIleSerProThraIa 280
Db 22068 CCAACCGCCCAAGGATTCGCGAGACGCTTCGCAACTTGACGAAATCTCCACCTGCG 22127
QY 281 TyrLeuThrValProIySgIyTTPGluGluLeuValGlyAlaLeuGluIArgAspSerThr 300
Db 22128 TACCTCACTGCGCAAGAGGTGGAGGAATTAATGAGGTGCTTGGACCGAGACATACC 22187
QY 301 LeuArgGluArgPheAlaArgMetLeuPhePhePheAlaAlaAlaGlyLeuSer 320
Db 22188 CTGGCGCAACGCTTCTTCGCTCGCATGAGCTGTTCTTCTTCGCGCGGCTGGGTTGCG 22247
QY 321 GInGlyIleTTPaPaArgLeuAspArgValAlaGluGlnHisCySgIyGluArgIleArg 340
Db 22248 CAAGGATCTGGGATCGTTTGACCGGGTCCGTGAACAGACTGTGGTGAACGCACTTCCG 22307
QY 341 MetMetAlaGlyLeuGlyMetThrGluThraIaProSerCySThrPheThrThGlyPro 360
Db 22308 ATGATGGCGGGTCTGGGCGATGACGAGACTGCTCTCTCTCTTACCAACCGGACCG 22367
QY 361 LeuSerMetAlaGlyTyrIleGlyLeuProAlaProGlyCySgIyValIleLeuValPro 380
Db 22368 CTGTCGATGGCTGCTTACATTTGGCTGCGACGCTGCTGCTGCTGCTGCTGCTGCTGCT 22427
QY 381 ValAspGlyIySLeuGluGlyArgPheHisGlyProHisValMetSerGlyTyrTTPArg 400
Db 22428 GTCCATGGAAATTTGAAAGGCGTTTCATGCTCGACGTCATGAGCGGCTTACGGGCT 22487
QY 401 AlaProGluGlnAsnAlaGlnAlaPheAspGluGluGlyTyrTyrCySAspGlyAspAla 420
Db 22488 GCTCTCTAACAATAATGCGCAAGCTTGCACAGAGAGGCTATTAATCTGCTCGGAGATGCC 22547
QY 421 IleYsLeuAlaAspProAlaAspProGlnIySgIyLeuMetPheAspGlyArgIleAla 440
Db 22548 ATCAAAATGGCGAGATCTGCGCATCTCAGAAAGCTCTGATGTTTGACGGTCGAAATGCT 22607
QY 441 GluAspPheIySLeuSerSerGlyValPheValSerValGlyProLeuArgThrArgAla 460
Db 22608 GAAAGCTTCAAGCTGCTCCTGAGGGGTATTTGTCAAGCGTGGGCGCATTCGCGAGCGG 22667
QY 461 ValLeuGluGlyIySLeuTyrValLeuAspValValValAlaAlaProAspArgGluCys 480
Db 22668 GTTCTGGAAGCGCGCTCTTACGCTCTGACCTAGTGTGTGCTGCTCTGATCTGAAATGC 22727
QY 481 LeuGlyLeuLeuValPheProArgLeuLeuAspCysArgAlaLeuSerGlyLeuGlyLys 500
Db 22728 CTTGGATTGCTCTGCTTCCGCGCTTCTGCACTGCGCTGCTGCTGCTGCTGCTGCTGCTG 22787
QY 501 GluAlaSerAspAlaGluValIleuAlaSerGluProValArgAlaTTPPheAlaAspTTP 520
Db 22788 GAGGCGTCGAGCGCGGAGTCTTGGCAGTGAAGCGGCTGGGCTGGGTTTGCTGACTGG 22847
QY 521 LeuIyAspArgLeuAsnArgGluAlaThrGlyAsnAlaSerArgIleMetTTPValGlyLeu 540
Db 22848 CTCAAAAGACTCATTCAGAAAGCAACTGGCAATCCAGTCCCATCATGTGGGTAGGGCTC 22907
QY 541 LeuAspThrProProSerIleAspIySgIyGluValThrAspIySgIySerIleAsnGln 560
Db 22908 CTGCAATCGCGCGCTGATTTGATTAAGGCGAGGCTGCTGCAAGAGGCTGATTAACGAG 22967
QY 561 ArgAlaValIleuGlnTTPArgSerAlaIySValAspAlaLeuTyrArgIyGluAspGln 580
Db 22968 CGGCGCTTTTTCGAATGCGCGTGGCGAAAGTTGATGCGGCTGATTCGTGGTGAAGATCAA 23027
QY 581 SerMetLeuArgAspGluAlaThrLeu 589
Db 23028 TCCATGCTGCGTGAAGAGCCACACTG 23054
```

```
RESULT 4
US-09-750-986D-1
; Sequence 1, Application US/09750986D
; Publication No. US20030228670A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Steinbuechel, Alexander
; APPLICANT: Priefert, Horst
; APPLICANT: Rabenhorst, Jurgen
; TITLE OF INVENTION: SYNTHETIC ENZYMES FOR THE PRODUCTION OF CONFERYL.
; TITLE OF INVENTION: ALCOHOL, CONFERYLALDEHYDE, FERULIC ACID, VANILLIN AND
; TITLE OF INVENTION: VANILLIC ACID AND THEIR USE
; FILE REFERENCE: Bayer-9998-CAO
; CURRENT APPLICATION NUMBER: US/09/750,986D
; CURRENT FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: 196 49 655.1 GERMANY
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO: 1
; LENGTH: 32679
; TYPE: DNA
; ORGANISM: Pseudomonas sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3146)..(3997)
; OTHER INFORMATION: gene = "ORF1"
US-09-750-986D-1

Alignment Scores:
Pred. No.: 0 Length: 32679
Score: 3082.00 Matches: 589
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0

US-09-750-986D-30 (1-589) x US-09-750-986D-1 (1-32679)
QY 1 MetArgSerLeuGluAlaLeuLeuProPheProGluIyArgIleLeuGluArgLeuGluHis 20
Db 21288 ATGGGTTCTCTCAGAGGGCTTCTTCTCCCGGGTGAATTTCTTGAAGCGCTTGACCAT 21347
QY 21 TTPAlaIySThrArgProGluGlnThrCySValAlaAlaArgAlaAlaAsnGlyGluTTP 40
Db 21348 TGGGCTAAGACCCGTCAGAAACAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 21407
QY 41 ArgArgIleSerTyrAlaGluMetPheHisAsnValArgAlaIleAlaGlnSerLeuLeu 60
Db 21408 CGTGTATCAGCTACGCGGAATGTTCCACAACGTCGCCCATCCGACAGAGCTTCTT 21467
QY 61 ProTyrGlyLeuSerAlaGluArgProLeuLeuIleIySLeuGlyAsnAspLeuGluHis 80
Db 21468 CTTTACGCACTATCGGACGAGCGCTGCTTATGCTCTCTGGAATGACTGGAAACAT 21527
QY 81 LeuGlnLeuAlaPheGlyAlaMetTyrAlaGlyIleProTyrCySProValSerProAla 100
Db 21528 CTTACGCTGCAATTTGGGCTATGTATGCGGCAATTCCTTATTCGCCGGTGTCTCTGCT 21587
QY 101 TyrSerLeuLeuSerGlnAspLeuAlaIySLeuArgHisIleValGlyLeuLeuGlnPro 120
Db 21588 TATTACTGCTGTCGCAAGATTTGGCGAAGCTGCTGACATCGTAGGCTTTCTGCAACCG 21647
QY 121 GlyLeuValPheAlaAlaAspAlaAlaProPheGlnArgAlaIleGluThrIleLeuPro 140
Db 21648 GAGCTGCTCTTCTGCGCATGAGCACCTTTCAAGGCCCAATTGAGCAACATTTGCGCG 21707
QY 141 AspAspValProAlaIlePheThrArgGlyGluLeuAlaGlyArgArgThrValSerPhe 160
Db 21708 GACGACGTCGCCGCAATCTTCACTCGAGGCAATTTGCCCGGGCGGCGACGAGTATTT 21767
QY 161 AspSerLeuLeuGluGlnProGlyIyIleGluAlaAspAsnAlaPheAlaAlaThrGly 180
Db 21768 GACAGCTGCTGGAGGAGCGCTGTGGGATTTGAAGCAGATATATGCTTTGGCGCACTGGC 21827
QY 181 ProAspThrIleAlaIySPhleLeuPheThrSerGlySerThrIySLeuPProIyAlaVal 200
Db 21828 CCCGATACGATTCGCAAGTTCTTGTTCACATTCTGCTCTACCAAACTGCTTAAGCGCGTG 21887
```

DB 1321 GAAGACTTCAAGCTGTCCTCAGGGGATATTGTCAAGCTTGGGCGATTGGCAGCGGGCG 1380
QY 461 ValLeuGluGlyGlySerTyrValLeuAspValValAlaAlaProAspArgGluCys 480
DB 1381 GTTCTGGAAAGCGCGCTCTTAACGTCTGGACGTAGTGTCTCTCTCTGATCGATGTC 1440
QY 481 LeuGlyLeuLeuValPheProArgLeuLeuAspCysArgAlaLeuSerGlyLeuGlyLeu 500
DB 1441 CTGGATTGCTGCTGTTTCCGCGCTCTTCTGCACTCCGCTCTTGTGCGGGCTAGGAAA 1500
QY 501 GluAlaSerAspAlaGluValLeuAlaSerGluProValArgAlaTrpPheAlaAspTrp 520
DB 1501 GAGCGCGTGGAGCCCGAGGCTGTTGCCAGTGAAGCGGCTTGGCTGTAAGTGG 1560
QY 521 LeuLeuArgLeuAspArgGluAlaThrGlyAspAlaSerArgGlyMetTrpValGlyLeu 540
DB 1561 CTCAACGACTCACTCAACAGAAAGCACTGGCAATGCCAGTGCATCAAGTGGGATGGCTC 1620
QY 541 LeuAspThrProProSerIleAspIleAspGlyGluValThrAspIleAspSerIleAspGln 560
DB 1621 CTGGATACCGCGCGCTGCTGATTGATTAAGGCGAGGCTCACTACAAAGGCTCGATCAACCA 1680
QY 561 ArgAlaValLeuGlnTrpArgSerAlaValAspAlaLeuTrpArgGlyGluAspGln 580
DB 1681 CGGCGCTGTTTGCATAGCGCGTGGCGAAAGTTGATGCGTGTATGTTGGAAGATCAA 1740
QY 581 SerMetLeuArgAspGluAlaThrLeu 589
DB 1741 TCCATGCTGCTGACGAGGCCACACTG 1767

RESULT 3

US-08-976-063C-1

Sequence 1, Application US/08976063C
Publication No. US20020182697A1

GENERAL INFORMATION:

APPLICANT: Alexander Steimbuchel; Horst Priefert; Jürgen Rabenhorst

TITLE OF INVENTION: SYNTHETIC ENZYMES FOR THE PRODUCTION OF

TITLE OF INVENTION: CONFERRAL ALCOHOL, CONFERRALDEHYDE, FERULIC ACID, VANILLIN A

NUMBER OF SEQUENCES: 42

CORRESPONDENCE ADDRESS:

ADDRESSEE: SPRUNG KRAMER SCHAEFER & BRISCOE

STREET: 660 White Plains Road

CITY: Tarrytown

STATE: New York

COUNTRY: U.S.A.

ZIP: 10591-5144

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB storage

COMPUTER: HP VECTRA

OPERATING SYSTEM: DOS

SOFTWARE: WordPerfect 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/976,063C

FILING DATE: 21-NOV-1997

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 196 49 655.1 (Germany)

FILING DATE: 29-NOV-1996

ATTORNEY/AGENT INFORMATION:

NAME: Kurt G. Briscoe

REGISTRATION NUMBER: 33,141

REFERENCE/DOCKET NUMBER: Bayer 9998-CAO

TELECOMMUNICATION INFORMATION:

TELEPHONE: (914) 332-1700

TELEFAX: (914) 332-1844

TELEX:

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 32679 base pairs

TYPE: Nucleic Acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (Genomic)
HYPOTHEtical: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Pseudomonas sp.
STRAIN: HR199
FEATURE:
NAME/KEY: CDS
LOCATION: 3146..3997
OTHER INFORMATION: /gene="ORF1"

Alignment Scores:

Score:	0	Length:	32679
Percent Similarity:	100.00%	Matches:	589
Best Local Similarity:	100.00%	Conservative:	0
Query Match:	100.00%	Mismatches:	0
DB:	8	Indels:	0
		Gaps:	0

US-09-750-986D-30 (1-589) x US-08-976-063C-1 (1-32679)

QY 1 MetArgSerLeuGluAlaLeuLeuProPheProGlyArgIleLeuGluValLeuGluLys 20
DB 21288 ATGCGTCTCTCGAGGCGCTTCCCTTCCGCGGTCGAATTCTTGAGCGTTCGAGCAT 21347
QY 21 TrpAlaIleThrArgProGluGlnThrCysValAlaAlaArgAlaAlaLeuGlyGluTrp 40
DB 21348 TGGGCTTAAGCCCGCTCCAGAAACAACTCGTGTGTCGACAGGGCGGCAATGGGAAATGG 21407
QY 41 ArgArgIleSerTyrAlaGluMetPheHisAsnValArgAlaIleAlaGlnSerLeuLeu 60
DB 21408 CGTCGATACAGTACCGGAAATGTTCCACAGTCCGGCCATGCCACAGAGCTTGCTT 21467
QY 61 ProTyrGlyLeuSerAlaGluArgProLeuLeuIleValSerGlyAsnAspLeuGluHis 80
DB 21468 CTTTACGGACTAATCGGACAGGCGCTCCGCTTATGCTCTGGAATGACCTCGAAACAT 21527
QY 81 LeuGlnLeuAlaPheGlyAlaMetTyrAlaGlyIleProTyrCysProValSerProAla 100
DB 21528 CTTCAGCTGCAATTTGGGCTATGATGCGGGAATTCCTATGCCCCGGTGTCTCTGCT 21587
QY 101 TyrSerLeuLeuSerGlnAspLeuAlaLysLeuArgHisAlaIleValGlyLeuGlnPro 120
DB 21588 TATTCATGCTGTCGCAAGATTTGGGAGCTGCCATCATGTAGTCTTCGCAACCG 21647
QY 121 GlyLeuValPheAlaAlaAspAlaAlaProPheGlnArgAlaIleGluThrIleLeuPro 140
DB 21648 GGAAGTGTCTTGTCTCCGATGACGACCTTTCACGCGCAATTGAGACCATTTGCGCG 21707
QY 141 AspAspValProAlaIlePheThrArgGlyGluLeuAlaGlyValArgTrpHisSerPhe 160
DB 21708 GAGAGCTGCGCCCAATCTTCACTCGAGGGAATTTGCGGGGCGCAACGAGGATTT 21767
QY 161 AspSerLeuLeuGlnProGlyIleGluAlaAspAsnAlaPheAlaAlaTrpGly 180
DB 21768 GACAGCTGCTGAGACGAGCTGTTGAGTGGAGCAGATATCTCTTGGGCAACTGGC 21827
QY 181 ProAspThrIleAlaLysPheLeuPheThrSerGlySerThrLysLeuProLysAlaVal 200
DB 21828 CCGATACGATTCGCAAGTCTTGTTCACCTTGGCTTACCAACTGAGCTTAAGGCGGTG 21887
QY 201 ProThrThrGlnArgMetLeuCysAlaAsnGlnGluMetLeuGlnThrPheProVal 220
DB 21888 CCGACTACTAGGGAATGCTTGGCCAAACAGCAATGTTTGCAAACTTCCCGGTT 21947
QY 221 PheGlyGluGluProProValLeuValAspTrpLeuProTrpAsnHisThrPheGlyGly 240
DB 21948 TTTGGTGAAGAGCGCGGCTGCTGAGTGGTTCGCGGAAACCACTTCCGCGCG 22007
QY 241 SerHisAsnIleGlyIleValIleValLysArgGlyGlyThrTyrTyrLeuAspAspGlyLys 260
DB 22008 AGCCACAACATCGGATCGTGTGTACAAACGCGGCGACGTAATCACTTGAAGCGGTAAA 22067

QY 521 LeuLYeArGLeuSArGjuaLaThrGlyASnAlaSerArgIleMetrPValGlyLeu 540
Db 1561 CTCAAAGCATCATGAGAGACACTGGCAATGCCAGTCAGTCATCATGTGGATAGGGCTC 1620
QY 541 LeuAspThrProProSerIleAspLYGlyGlyValaThrAspLYSerIleAsnGln 560
Db 1621 CTCGATACGCCGCCGCTCGATTGATTAAGGGCGAGTCACTGACAAAGGCTCGATCAACACG 1680
QY 561 ArgAlaValLeuGlnTrpArgSerAlaValaAspAlaLeuTrpArgGlyValaAspGln 580
Db 1681 CGGGCTGTTTGCAATGGCCGCTGGCGCAAAAGTTGATGCGCTGATCGTGAAGATCAA 1740
QY 581 SerMetLeuArgAspGluAlaThrLeu 589
Db 1741 TCATGCTGCTGACGAGGCGCACACTG 1767
RESULT 2
US-09-750-986D-29
; Sequence 29, Application US/09750986D
; Publication No. US20030228670A1
; GENERAL INFORMATION:
; APPLICANT: Steinbuechel, Alexander
; APPLICANT: Pfelefert, Horst
; APPLICANT: Rabenhorst, Jurgen
; TITLE OF INVENTION: SYNTHETIC ENZYMES FOR THE PRODUCTION OF CONIFERYL
; TITLE OF INVENTION: ALCOHOL, CONIFERYLALDEHYDE, FERULIC ACID, VANILLIN AND
; TITLE OF INVENTION: VANILLIC ACID AND THEIR USE
; FILE REFERENCE: Bayer-9998-CNO
; CURRENT APPLICATION NUMBER: US/09/750, 986D
; PRIOR FILING DATE: 2000-12-28
; PRIOR FILING DATE: 1996-11-29
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 29
; LENGTH: 1770
; TYPE: DNA
; ORGANISM: Pseudomonas sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1767)
; OTHER INFORMATION: product = "Ferulasaure-CoA-Synthetase" / gene =
; OTHER INFORMATION: "fcs"
US-09-750-986D-29
Alignment Scores:
Pred. No.: 0 Length: 1770
Score: 3082.00 Matches: 589
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0
US-09-750-986D-30 (1-589) x US-09-750-986D-29 (1-1770)
QY 1 MetArgSerLeuGlnAlaLeuLeuProPheProGlyValArgIleLeuGlnArgLeuGlnHis 20
Db 1 ATCGCTTCTTCAGAGCGCTTCTTCCTCCGGGTCGAATTTCTTGAGCTTCAGACAT 60
QY 21 TrpAlaLeuThrArgProGlnGlnThrCysValaAlaAlaArgAlaAlaAsnGlyGluTrp 40
Db 61 TGGGCTTAAGACCCGCTCCAGAAACAACCTGCTGCTCCAGAGCGCGCAATGGGGAATGG 120
QY 41 ArgArgIleSerTrpAlaGluMetPheHisAsnValaArgAlaIleAlaGlnSerLeuLeu 60
Db 121 CGTCGATACGCTACGCGGAAATGTTCCACACGTCGCGCCATCGCACAGACTTGCTT 180
QY 61 ProTrpGlyLeuSerAlaGluArgProLeuLeuIleValSerGlyAsnAspLeuGlnHis 80
Db 181 CTCTTAGGAGATATCGGACAGGCGTCCGCTCTATGCTCTCTGAAATGACTCTGGAACAT 240
QY 81 LeuGlnLeuAlaPheGlyAlaMetTrpAlaGlyIleProTrpCysProValSerProAla 100

Db 241 CTTCAGCTGCATTTGGGGGCTATGTATGCCGGCATTTCCCATTTGCCGGGTCTCCGCT 300
QY 101 TrpSerLeuLeuSerGlnAspLeuAlaValLeuArgHisIleValaGlyLeuLeuGlnPro 120
Db 301 TATTCACTGCTGTCGCAAGATTTGGCGAACTCTGTCACATCTGATGCTTCTGCAACCG 360
QY 121 GlyLeuValPheAlaAlaAspAlaAlaProPheGlnArgAlaIleGluThrIleLeuPro 140
Db 361 GAGCTGGTCTTCTGCTCCGATGACAGCACTTTCAGAGGCGCAATTGAGACCATTTGCGC 420
QY 141 AspAspValProAlaIlePheThrArgGlyGlyLeuAlaGlyValArgThrValSerPhe 160
Db 421 GACGAGCTGCCCAATCTTCACCTCGAGGAAATTGGCCGGGCGGCACAGGTAAGTTT 480
QY 161 AspSerLeuLeuGlnProGlyIleGlyAlaAspAsnAlaPheAlaAlaThrGly 180
Db 481 GACAGCTGCTGAGAGACCTGGTGGAATTGAAGCAATATGCTTTGGCGCACTGGC 540
QY 181 ProAspThrIleAlaValPheLeuPheThrSerGlySerThrLYsLeuProLYsAlaVal 200
Db 541 CCGGATACGATTCGCAAGTTCTTGTCACTTGGCTCTACCAACTGCTTAAGGGGCTG 600
QY 201 ProThrThrGlnArgMetLeuCysAlaAsnGlnGlnMetLeuLeuGlnThrPheProVal 220
Db 601 CCGACTACTCAGCGAATGCTTCGCCCAATCACAGATGCTTCGCAAACTTCCGCTT 660
QY 221 PheGlyGlnGluProProValaIleValaAspTrpLeuProThrPheGlyGly 240
Db 661 TTGGTGAAGAGCGCGGCTGCTGTGGATGCTGTTGCGTGAACACACACTTTCGGCGGC 720
QY 241 SerHisAsnIleGlyIleValaLeuTrpAsnGlyGlyThrTrpTrpLeuAspAspGlyLYs 260
Db 721 AGCCACAAATCGCGCATCGGTGTACAAGCGGCGACGATACCTTGACGACGCGTAA 780
QY 261 ProThrAlaGlnGlyPheAlaGluThrLeuArgAsnLeuSerGluIleSerProThrAla 280
Db 781 CCAACCGCCCMAGGGTTCGCCGAGACGCTTCGCAACTTGAGCCAAATCTCTCCACATGCG 840
QY 281 TrpLeuThrValProLYsGlyTrpGluGluValaGlyAlaLeuGlnArgAspSerThr 300
Db 841 TACCTCACTGTGCGAAGAGCTGGAGAAATTGAGGTGCTTGGACGACAGATACC 900
QY 301 LeuArgLYsArgPhePheAlaArgMetLYsLeuPhePhePheAlaAlaArgIleLeuSer 320
Db 901 CTGCGGGAACGCTTCTTCGCTGCGATGAAGCTGTTCTTCTTCGCGGGCTGGCTGTG 960
QY 321 GlnGlyIleTrpAspArgLeuAspArgValaAlaGluGlnHisCysGlyGlyValaGlyIleArg 340
Db 961 CAAGGATCTGGGATGTTTGGACCGGGTCGCTGAACAGCATGTGGTGAGCCCATTCGC 1020
QY 341 MetMetAlaGlyLeuGlnMetThrGluThrAlaProSerCysThrPheThrTrpGlyPro 360
Db 1021 ATGATGCGGGGTGTGGACATGACGAGACTGCTCTTCCTTCGCACTTTTACACGGAACG 1080
QY 361 LeuSerMetAlaGlyTrpIleGlyLeuProAlaProGlyCysGluValLYsLeuValPro 380
Db 1081 CTGTGATGGCTGTGATACATTTGGGCTGCCAGCGCTGGCGGAGGCAAGCTCGTTCCG 1140
QY 381 ValaAspGlyLYsLeuGlnGlyValArgPheHisGlyProHisValaMetSerGlyTrpTrpArg 400
Db 1141 GTCGATGGGAAATTTGAAAGGCGTTTCCATGTGCTCCACCTCATGAGCGGCTCTGGCGT 1200
QY 401 AlaProGluGlnAsnAlaGlnAlaPheAspGluGluGlyTrpTrpCysSerGlyAspAla 420
Db 1201 GCTCTGAAACAAATGCTCCAGGCTTTCAGAGGAAAGGCTATTATCTGCTCCGGTATGCC 1260
QY 421 IleLYsLeuAlaAspProAlaAspProGlnLYsGlyLeuMetCysPheAspGlyArgIleAla 440
Db 1261 ATCAAAATTGGCAATCTCTGCGATCTCTCAAGAAAGCTGATGTTTACCGGTGAATTGCT 1320
QY 441 GluAspPheLYsLeuSerSerGlyValaPheValSerValaGlyProLeuArgThrArgAla 460

SOFTWARE: wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: us/08/976, 063C
FILING DATE: 21-NOV-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 196 49 655.1 (Germany)
FILING DATE: 29-NOV-1996
ATTORNEY/AGENT INFORMATION:
NAME: Kurt G. Briscoe
REGISTRATION NUMBER: 33,141
REFERENCE/DOCKET NUMBER: Bayer 9998-CAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (914) 332-1700
TELEFAX: (914) 332-1844
TELEX:
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 1770 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1767
OTHER INFORMATION: /product=
OTHER INFORMATION: "Ferulaseure-CoA-Synthetase"
OTHER INFORMATION: /gene="fcs"
US-08-976-063C-29

Alignment Scores:
Pred. No.: 0 Length: 1770
Score: 3082.00 Matches: 589
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0

US-09-750-986D-30 (1-589) x US-08-976-063C-29 (1-1770)

Qy 1 MetAArgSerLeuGluAlaLeuLeuProPheProGluYArgIleLeuGluArgLeuGluHis 20
Db 1 ATGGGTCCTCTCGAAGCGCTTCTCCCTCCGGGTGAATCTTGAGCGCTCGACAT 60
Qy 21 TTPAlaYerThrArgProGluGlnThrCysValAlaAlaArgAlaAlaAngIYgIuTrp 40
Db 61 TGGGCTAAGACCCCTCCAGAACAACTCGCTTCCTCCAGGCGGCAATGGGGAATGG 120
Qy 41 ArgAArgIleSerYArgAlaGluMetPheHisValAlaArgAlaAlaAngIleu 60
Db 121 CGTCGTTCAGCTACCGCGAAATGTTCCACAACTCCGCCCATCGCACAGAGCTTCTT 180
Qy 61 ProTYrGlyLeuSerAlaGluArgProLeuLeuIleValSerGlyAsnAspLeuGluHis 80
Db 181 CTTTACGAGCTATCGGAGAGCGTCGCTGTATCGTCTTGAAATGACCTGAAACAT 240
Qy 81 LeuGlnLeuAlaPheGlyAlaMetYrAlaGlyIleProTYrCysProValSerProAla 100
Db 241 CTTTACGCTGATTTGGGGCTATGTATCGGGGCAATCCCTATTCGCGGTCTCTGCT 300
Qy 101 TyrSerLeuLeuSerGlnAspLeuAlaLysLeuArgHisIleValGlyLeuLeuGlnPro 120
Db 301 TATTTCATGCTGTTCGCAAGATTGGCGAAGCTGCTACATCGTAGGTCCTTCTGCAACG 360
Qy 121 GlyLeuValAlaPheAlaAlaAspAlaAlaProPheGlnArgAlaIleGluTrpIleLeuPro 140
Db 361 GGACTGCTCTTGTGCGAGTGCAGACCTTTCAGAGCGGCAATTGAGACCATTCGCG 420
Qy 141 AspAspValProAlaIlePheThrArgGlyGluLeuAlaGlyArgArgThrValSerPhe 160

Db 421 GACGAGCTGCCCGCAATCTTCACTCGAGGCGAATTGGCGGGCGCGCACGCTGACTTT 480
Qy 161 AspSerLeuLeuGluGlnProGlyGlyIleGluAlaAspAsnAlaPheAlaAlaThrGly 180
Db 481 GACAGCTCTGGAGCAGCTGGGTGGATTGAGCAGATTAATGCTTTGGCGCACTGGC 540
Qy 181 ProAspThrIleAlaLysPheLeuPheThrSerGlySerThrIlysLeuProIysAlaVal 200
Db 541 CCGGATACGATTGCCAAGTTCTTTGTTCACTTGCGCTTACCAAACTGCTTAAGCGGTG 600
Qy 201 ProThrThrGlnArgMetLeuCysAlaAsnGlnMetLeuLeuGlnThrPheProVal 220
Db 601 CCACACTACTCAGGAAATGCTCGCGCAATCAGCAATCTTCTGCAACTTCCCGGTT 660
Qy 221 PheGlyGluGlnProProValLeuValAspTrpLeuProTrpAsnHisThrPheGly 240
Db 661 TTGGTGAAGAGCGCGGCTGCTGGAGCTGGTTCCTTGAAACACACTTTCGCGGCG 720
Qy 241 SerHisAsnIleGlyIleValLeuTYrAangIYgIYThrTYrLeuAspAspGlyLys 260
Db 721 AGCCACAATCGGATCGGTGTTCACACGCGGCGACGTACTACCTTGACGACGGTAA 780
Qy 261 ProThrAlaGlnGlyPheAlaGluThrLeuArgAsnLeuSerGluIleSerProThrAla 280
Db 781 CCAACCGCCCAAGGGTTCCCGCAGACGCTTCGCACTTAGCGCAATCTTCCACTGCG 840
Qy 281 TyrLeuThrValProIysGlyTrpGluGluLeuValGlyAlaLeuGluArgAspSerThr 300
Db 841 TACCTCACTGTGCGAAGCTGGAGAAATTAATGAGGACCTTGAAGCAGACATACC 900
Qy 301 LeuArgGluArgPhePheAlaArgMetLysLeuPhePhePheAlaAlaAlaGlyLeuSer 320
Db 901 CTGCGGGAACGCTTCTTCGCTGCATGAAGCTGTTCTTCTCGCGCGGTGGTGTGCG 960
Qy 321 GlnGlyIleTrpAspArgLeuAspArgValAlaGluGlnHisCysGlyGlyIuArgIleArg 340
Db 961 CAAGGATCTGGGATGTTTGAGCCGGTTCGCGTGAACACACATGTGTGAGCCCATTCG 1020
Qy 341 MetMetAlaGlyLeuGlyMetThrGluThrAlaProSerCysThrPheThrArgIlePro 360
Db 1021 ATGATGCGGGGTGGGACATGAGAGACTGCTTCTTCTGCACTTTTACACCGGACCG 1080
Qy 361 LeuSerMetAlaGlyTYrIleGlyLeuProAlaProGlyCysGlyValIysLeuValPro 380
Db 1081 CTGTGATGCTGCTGTACATTTGGGCTGCCAGCGCTGGCGGAGGTCAAGCTCGTTCCG 1140
Qy 381 ValAspGlyLysLeuGluGlyArgPheHisGlyProHisValMetSerGlyTYrTrpArg 400
Db 1141 GTCGATGGGAATGGAGAGGCTTTCCATGTGCCACGTCATGAGCGGCTTCTGCGCT 1200
Qy 401 AlaProGluGlnAsnAlaGlnAlaPheAspGluGluGlyTYrTYrCysSerGlyAspAla 420
Db 1201 GCTCTGAACAAATGCCCAAGGCTTCGACGAGAAAGGCTATTAAGCTCCGGATGCC 1260
Qy 421 IleLysLeuAlaAspProAlaAspProGlnLysGlyLeuMetPheAspGlyArgIleAla 440
Db 1261 ATCAATTTGGCAATCTCTGCGATCTCCAGAAAGCTGATGTTTGAACGCTCAATTGCT 1320
Qy 441 GluAspPheLysLeuSerSerGlyValPheValSerValGlyProLeuArgThrArgAla 460
Db 1321 GAAGACTTCAAGCTGCTTCAGGGGTATTTGTACAGGTTGGGCGCATTTGCCACGCGG 1380
Qy 461 ValLeuGluGlyLysSerTYrValLeuAspValValAlaAlaProAspArgLysCys 480
Db 1381 GTTCTGAAGCGCGCTTTCATGCTTCGAGAGTATGATGCTCTCTGATCTGAAATGC 1440
Qy 481 LeuGlyLeuLeuValPheProArgLeuLeuAspCysArgAlaLeuSerGlyLeuGlyLys 500
Db 1441 CTTGGAATGCTGCTGTTTCCGCGCTTCTCTGATGCGGCTTGTGCGGGGCTAGGAAA 1500
Qy 501 GluAlaSerAspAlaGluValAlaAlaSerGluProValAlaArgAlaTrpPheAlaAspTrp 520
Db 1501 GAGGCGTCGAGCCGAGGCTTGTGCAATGACCGGCTTCGGGCTGTTGTGCTGACTGG 1560

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 8, 2005, 02:03:17 ; Search time 791 Seconds
(without alignments)
4278.544 Million cell updates/sec

Title: US-09-750-986D-30
Perfect score: 3082
Sequence: 1 MRSLEALPPGRILRLER.....KVDALYRGDSQMLRDEATL 589

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4300275 seqs, 2872944193 residues

Total number of hits satisfying chosen parameters: 8600550

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=x1p
-O/cg2n1/USPTO.epool.p/US9750986/runat_03022005_071745_18678/app_query.fasta.1.775
-DB=Published Applications_NA -QFMT=fastap -SUFFIX=mpd -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blcsum62
-TRANS=human40.cdt -LIST=45 -DOCCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09750986 @CGN 1.1 723 @runat_03022005_071745_18678
-NCPU=6 -ICPU=3 -NO_MMAR -LARGQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -18678
-LONGLOG -DEV.TIMEOUT=120 -MARR.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications_NA:*
1: /cg2n2_6/prodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cg2n2_6/prodata/2/pubpna/PCRT_NEW_PUB.seq:*
3: /cg2n2_6/prodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cg2n2_6/prodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cg2n2_6/prodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cg2n2_6/prodata/2/pubpna/PCRTUS_PUBCOMB.seq:*
7: /cg2n2_6/prodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cg2n2_6/prodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cg2n2_6/prodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cg2n2_6/prodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cg2n2_6/prodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cg2n2_6/prodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cg2n2_6/prodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cg2n2_6/prodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cg2n2_6/prodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cg2n2_6/prodata/2/pubpna/US10E_PUBCOMB.seq:*
17: /cg2n2_6/prodata/2/pubpna/US10F_PUBCOMB.seq:*
18: /cg2n2_6/prodata/2/pubpna/US10G_NEW_PUB.seq:*
19: /cg2n2_6/prodata/2/pubpna/US11_NEW_PUB.seq:*
20: /cg2n2_6/prodata/2/pubpna/US60_NEW_PUB.seq:*
21: /cg2n2_6/prodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3082	100.0	1770	8	US-08-976-063C-29
2	3082	100.0	1770	8	US-09-750-986D-29
3	3082	100.0	32679	8	US-08-976-063C-1
4	3082	100.0	32679	10	US-09-750-986D-1
5	1457	47.3	1725	15	US-10-369-493-44177
6	1294	42.0	1728	15	US-10-369-493-35384
7	1294	42.0	1728	15	US-10-369-493-38427
8	1294	42.0	1728	15	US-10-369-493-38903
9	1226	39.8	1660	15	US-10-369-493-38007
10	972	31.5	1866	15	US-10-369-493-40734
11	383.5	12.4	1632	15	US-10-369-493-40734
12	335.5	10.9	1908	15	US-10-369-493-23931
13	335.5	10.9	9025608	15	US-10-156-761-5544
14	324	10.5	1746	15	US-10-369-493-34073
15	314.5	10.2	2256646	17	US-10-470-565-1
16	286.5	9.3	1686	15	US-10-369-493-37423
17	284	9.2	2103	14	US-10-119-136-124
18	283	9.2	1830121	14	US-10-329-960-1
19	283	9.2	1830121	16	US-10-329-670-1
20	283	9.2	1830121	18	US-10-158-865-1
21	283	9.2	1830121	15	US-10-369-493-23998
22	281	9.1	1839	15	US-10-369-493-43422
23	281	9.1	1839	15	US-10-369-493-43422
24	279	9.1	2103	10	US-09-906-419-6
25	279	9.1	2103	14	US-10-119-136-6
26	279	9.1	2103	16	US-10-410-031-6
27	278.5	9.0	1839	15	US-10-369-493-44573
28	278	9.0	2106	10	US-09-906-419-5
29	277.5	9.0	1515	15	US-10-369-493-35776
30	277.5	9.0	1794	9	US-10-369-493-31796
31	277	9.0	2097	15	US-09-938-842A-2269
32	277	9.0	2097	11	US-09-938-842A-2269
33	275	8.9	2257	14	US-10-119-136-5
34	275	8.9	2257	16	US-10-410-031-5
35	273	8.9	2287	18	US-10-739-930-1310
36	272	8.8	1767	15	US-10-369-493-44932
37	268	8.7	1467	15	US-10-369-493-44734
38	268	8.7	1500	15	US-10-369-493-43198
39	265.5	8.6	85632	16	US-10-461-194-1
40	265	8.6	1806	15	US-10-369-493-24610
41	265	8.6	1914	15	US-10-156-761-4802
42	264.5	8.6	1639	10	US-09-838-469-10
43	264.5	8.6	2542	18	US-10-425-115-115985
44	264.5	8.6	9025608	15	US-10-156-761-1
45	263.5	8.5	1639	10	US-09-813-279B-8

ALIGNMENTS

RESULT 1
US-08-976-063C-29
; Sequence 29, Application US/08976063C
; Publication No. US20020182697A1
GENERAL INFORMATION:
; APPLICANT: Alexander Steinhuechel; Horst Prieffert; Jurgen Rabenhorst
; TITLE OF INVENTION: SYNTHETIC ENZYMES FOR THE PRODUCTION OF
; TITLE OF INVENTION: CONFERYL ALCOHOL, CONFERYLALDEHYDE, FERULIC ACID, VANILLIN AN
; NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESSES:
; ADDRESSEE: SPRING KRAMER SCHAEFER & BRISCOE
; STREET: 660 White Plains Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10591-5144
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB storage
; COMPUTER: HP VECTOR
; OPERATING SYSTEM: DOS

This Page Blank (uspto)

```

; EARLIER APPLICATION NUMBER: PCT/US98/19494
; EARLIER FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: US 60/059,379
; EARLIER FILING DATE: 1997-09-19
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 42
; LENGTH: 1639
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Mutant Luciferase
; US-09-396-154-42

```

```

Alignment Scores:
Pred. No.: 1 08e-18 Length: 1639
Score: 258.00 Matches: 127
Percent Similarity: 36.52% Conservative: 79
Best Local Similarity: 22.52% Mismatches: 218
Query Match: 8.37% Indels: 141
Gaps: 23

```

US-09-750-986D-30 (1-589) x US-09-396-154-42 (1-1639)

```

Qy 30 CyValAlaAlaArgAlaAlaAsnGlyLutRpaRgaRglSerTyraAlaGluMetPhe 49
    |||::|||
Db 119 TGcATACATTCGAAATGCTCATACAAA---GAAATGTTTATATGAGAGATTCTG 175
Qy 50 HisAsnValArgAlaAlaGlnSerLeuLeuProTyrglyLeuSerAlaGluArgPro 69
    ::|||::|||
Db 176 AAACGTGCTGTCGTTAGCGGAAAGTTTMAAAAGATGATTMAAACAAACAGACACA 235
Qy 70 LeuLeuValSerGlyAsnAspLeuGluHisLeuGlnLeuAlaPheGlyAlaMetTyr 89
    ::|||::|||
Db 236 ATACCGGTGTGTACCGAAATAGCTCTCAATTTTCTCTCGTAATGATCATCTGTAT 295
Qy 90 AlaGlyIleProTyrcysProValSerProAlaTySerLeuLeuSerGlnAspLeuAla 109
    |||::|||
Db 296 CTTCGATTAATGTGGCAGCTTAAACGATTAATAC-----ATTGAACGTGAATTATA 349
Qy 110 LysLeuArgHisIleValGlyLeuLeuGlnProGlyLeuValPheAlaAlaAspAlaAla 129
    |||::|||
Db 350 -----CACAGCTCTGTGATTGTATTAACACGACGATAGTTTCTCCACAGAAATAC 400
Qy 130 ProPheGlnArgAlaIleGluThrIleLeuProAspAspValProAlaIlePheThrArg 149
    |||::|||
Db 401 ---TTCAAAAAGTACTGAAT-----GTAAAA 424
Qy 150 GlyLeuLeuAlaGlyArgThrValSerPheAspSerLeuLeuGlnGlnProGlyGly 169
    ::|||::|||
Db 425 TCTAATTTAAATCTATTTGAACATATATATATTTAGACTTAATGATGACTTAGAGGT 484
Qy 170 IleGluAlaAspAsnAlaPheAlaAlaThrGlyProAspThr----- 183
    ::|||::|||
Db 485 TATCAATGCGTCACACAACTTATTTCTCAAAATTCGATGTAATCTGACGTAAAAAA 544
Qy 184 -----IleAlaLysPheLeuPheThrSerGly 192
    ::|||::|||
Db 545 TTTTAACCAATATTTTATTCAGACAGATCAGGTGCTGTTGATTAATGTTTCTTCGCT 604
Qy 193 SerThrLysLeuProLysAlaValProThrThrGlnArgMetLeuCyAlaAsnGlnGln 212
    ::|||::|||
Db 605 ACAACTGTGTGCGGAAGGAGTCACTGTAACACAAAGATATGTTGCA-----CGA 658
Qy 213 MetLeuGlnThrPheProValPheGly-----GluGluProProValLeuVal 229
    ::|||::|||
Db 659 TTTTCTATTGCAAAAAGATCTTACTTTTGTATCCAACTTAATCCACGTACGCAATTTTA 718
Qy 230 AspThrLeuProThrPasnHisThrPheGlyGlySerHisAsnIleGly----- 245
    ::|||::|||
Db 719 ACGGTAAATCTTCCACCATGTTTGGTATGATGACACCATTAAGATTAATCTTACTTGT 778
Qy 246 -----IleValLeuTyraAsnGlyGlyThrTyrrTyrrLeuAspAspGlyLysProThr 262

```

```

Db 779 GGATTCGACTTGTCTTAATGCAC----- 802
Qy 263 AlaGlnGlyPheAlaGluThr-----LeuArgAsnLeuSerGluIleSerProThrAla 280
    |||::|||
Db 803 -----ACGTTGAGAGAAAACTATTTCTACATTCATTAACAATTAAGCGAAAGT 856
Qy 281 TyrLeuThrValProLysGlyTyrGluGluLeuValGlyAlaLeuGluArgAspSerThr 300
    |||::|||
Db 857 ACTTACTGTACCA-----ACATTAATGCGCATTTCTTGCAAAAAGTCACTTA 904
Qy 301 LeuArgGluArgPhePheAlaArgMetLysLeuPhePheAlaAlaAlaGlyLeuSer 320
    ::|||::|||
Db 905 GTTGAAAGATGACATTTATGACATTAAGAAATTCATCTGTGCGCACCTTATCA 964
Qy 321 GlnGlyIleThrAspArgLeuAspArgValAlaGluGlnHisCysGlyGluArgIleArg 340
    ::|||::|||
Db 965 AAAGAAATTT-----MetMetAlaGlyLeuGlyMetThrGluThrAlaProSer 353
Qy 341 -----MetMetAlaGlyLeuGlyMetThrGluThrAlaProSer 353
    ::|||::|||
Db 989 AAACGTTTAATTAATTAACCTTGTACGCGAAGGATGATTAACAGAAACCACTTCGCT 1048
Qy 354 CysThrPheThrThrGlyProLeuSerMetAlaGlyTyrIleGlyLeuProAlaProGly 373
    |||::|||
Db 1049 GTTTTAATTAACCGCAAGGTGACGCCAAACCGGATCAACGTGTAATAGTACATTT 1108
Qy 374 CysGluValLysLeuVal---ProValAspGlyLysLeuGluGly----- 387
    |||::|||
Db 1109 CAGCGTTTAAAGTTGTCATCTTACACAGGAAAAATTTGGGCGCAATGAACCTGCA 1168
Qy 388 -----ArgPheHisGlyProHisValMetSerGlyTyrTrpArgAlaProGluGlnAsn 405
    |||::|||
Db 1169 GAATGTATTTTAAAGCCCGCATGATTAAGAGGTATTAATATGAAGAACTACT 1228
Qy 406 AlaGlnAlaPheAspGluGlyTyrTyrCysSerGlyAspAlaIleLysLeuAlaAsp 425
    |||::|||
Db 1229 AAAGCAATTAATGATATATGACGATGTTGCGCTCTGCTGAT---ATTGCTTATTAAGC 1285
Qy 426 ProAlaAspProGlnLysGlyLeuMetPheAspGlyValArgIleAlaGluAspPheLeu 445
    |||::|||
Db 1286 AAT-----GATGGCCATTTTATATTTGTCGACAGCGCTG 1318
Qy 446 SerSerGlyVal-----PheValSerValGlyProLeuArgThrArgAlaValLeu 462
    |||::|||
Db 1319 AAGTCACTGATTAATATTAAGTTATCAGGTGACCTCGTGAATTTAGCGGAATCTC 1378
Qy 463 GluGlyGlySerTyrrValLeuAspValValAlaAlaProAspAspGluCysLeuGly 482
    |||::|||
Db 1379 TTACAAACATCCGTATATATTGTTGATGCCGCGCTTACTGTATACCGATGAAGCGCGGC 1438
Qy 483 LeuLeuValPheProArgLeuLeuAspCysArgAlaLeuSerGlyLeuGlyLysGluAla 502
    |||::|||
Db 1439 -----GAGCTTCC-AGCTGCAAGGTGTGTGT----- 1464
Qy 503 SerAspAlaGluValLeuAlaSerGluProValArgAlaIlePheAlaAspTyr----- 520
    |||::|||
Db 1465 -----ACAGACTGGAAATA 1479
Qy 521 LeuLysArgLeuAsnArg-----GluAlaThrGlyAlaAlaSerArgIle 535
    |||::|||
Db 1480 TCTTAACGAACAATCTTACAGATTAATGTTGCCAGATCAAGTTCACACGCCAAATGCGT 1539
Qy 536 MetTrpValGly 539
    |||::|||
Db 1540 ACGTGTGGGGGT 1551

```

Search completed: February 8, 2005, 04:55:18
 Job time : 7107 secs


```

Qy      2 ArgSerLeuGluAlaLeu-----ProPheProGlyArgIleLeu 15
Db      2448137 AGGTCAAGTAGGTGTTATCGTCGAGATTAGCGTCCCGCCATTTCATCTGCGC 2448196
Qy      16 GluArg-----LeuGluHisTrpAlaLeuSerProGluGln 28
Db      2448197 GACGACGACAAAGTCGGCGGACATGTTGTGAGCAT---GACGTCAGCATCCGCACTAC 2448253
Qy      29 ThrCysValAlaAlaArgAlaAlaSerGlyLeuTrpArgArgIleSerTrpAlaGluMet 48
Db      2448254 GTC---ATCTATCAACCGCTGATCGAGCGGCTCGGACCGCATGTCAGTGGCGAGGCA 2448310
Qy      49 PheHisAsnValArgAlaIleAlaGlnSerLeuLeuProTrpGlyLeuSerAlaGluArg 68
Db      2448311 GCCAACCAAGATGTCGCGCGGCTCTCGGTTGATTCATCTGGGGGTCAGCGCGCGCAT 2448370
Qy      69 ProLeuLeuIleValSerGlyAsnAspLeuGluHisLeuGlnLeuAlaPheGlyAlaMet 88
Db      2448371 CGGGTAGTCATCTTCTGCGCACCGGCTACAGAGTGGCGCATCTCGATTTCGGCATTCG 2448430
Qy      89 TrpAlaGlyIleProTrpCysProValSerProAlaTrpSerLeuLeuSerGlnAspLeu 108
Db      2448431 GCTGTGGGT-----CGGTCACCGTACCGCATCGACGACTCTGTCAGCGAG--- 2448478
Qy      109 AlaLeuLeuArgHisIleValGlyLeuLeuGlnProGlyLeuValPheAlaAlaAspAla 128
Db      2448479 ---CAGTGGCGCTGGGTTTACAAAGACTCCGAAAGCGGTGTTGTTGCGGAAACGAC 2448535
Qy      129 AlaProPheGlnArgAlaIleGluTrpIleLeuProAspAspValProAlaIlePheThr 148
Db      2448536 TCACAGCGCAATGTCGCGGCA---CTCTCGGCAAGCGTGGCGCGCTGCGGAG 2448589
Qy      149 ArgGlyLeuLeuAlaGlyArgArgThrValSerPheAspSerLeuLeuGlnProGly 168
Db      2448590 GATCTGAGATCGCGGCTTCGCGTCCCAAGCGCTCATCGGCTCAGCGAGGCGGCGCC 2448649
Qy      169 GlyIleGluAlaAspAsn-----AlaPheAlaIleThrGlyProAsp 182
Db      2448650 TCGGTGACCGCGCGGAGTACCGCGCGCTGCGGCACTACGCTGACGAGACCGCGC 2448709
Qy      183 ThrIleAlaLeuPheLeuPheThrSerGlySerThrIleLeuProValAlaValProThr 202
Db      2448710 ACGCTT-----ATCTACACTCGGGACACGACCGAGACCGCAAGGCGTGCAGTTG 2448760
Qy      203 ThrGlnArgMetLeuCysAlaAsnGlnMetLeuLeuGlnThrPheProValPheGly 222
Db      2448761 ACCCAATCCAAACCTGGTTCAAGATTAAAGCGCGCGCATATCAACCGCGCTGCTG 2448820
Qy      223 GluGluProProValLeuValAspTrpLeuProTrpAsnHisThrPheGlyGlySerHis 242
Db      2448821 CGCAAGGCTAGCGCGCTGCTGTTCTCTGCGCTAGCTATGCTG-----GCCGC 2448874
Qy      243 AsnIleGlyIleValLeuTrpAsnGlyGlyThrTrpTrpLeuAspArgIleYsProThr 262
Db      2448875 GCGATCATGATGCGCGCTTCCAC-----TCCAAAGTCAAC 2448910
Qy      263 AlGlnGlyPheAlaGluTrpLeuArgAsn-----LeuSerGluIleSerPro 278
Db      2448911 GTG---GGATTCAACGACGATCAAGATCTGCTGCGATGGTGGGAGTTTCAAACCG 2448967
Qy      279 ThrAlaTrpLeuThrValProValGlyTrpGluGluLeuValGlyAlaLeuGluArgAsp 298
Db      2448968 ACGGTGTGTGTGTGCGTCCGAGGAGTTCGAGAAAGTGTACAAACCGCGAGCAAGAC 2449027
Qy      299 SerThrLeuArgGluArgPhePheAlaArgMetLeuLeuPhePheAlaAla----- 316
Db      2449028 GCCGCCAACCC-----GCCAAAGGCGAATCTTCGAGATCGCGCGCAAGC 2449075
Qy      317 -----AlaGlyLeu-----Ser 320
Db      2449076 GCGGTGACTGAGGAGAACTTGGACCGCGCGGACCGGAGCTGACTTACGCGCGCAAG 2449135
Qy      321 GlnGlyIleTrpAspArgLeu----- 327

```

```

Db      2449136 CACCGGTGTTCACCGCGTGTGTCTACCGCAAGCTGCTCGGCACTGCTGGCAACTGC 2449195
Qy      328 -----AspArgValAlaGluGlnHisCysGly 336
Db      2449196 CGCGCGCGCTCTCCGCGCGCGCGCGCTGAGTGGCGCGGCTGTGTCATCTTATCCGCGC 2449255
Qy      337 GluArgIleArgMetLeuAlaGlyLeuGlyMetThrGluThrAlaProSerCysThrPhe 356
Db      2449256 GCGGTCTCACCATCTACGAGGATACCGCTGAGCGGACGATGGGGCGTCCGCATC 2449315
Qy      357 ThrThrGlyProLeuSerMetAlaGlyTrpIleGlyLeuProAlaProGlyCysGluVal 376
Db      2449316 AGCCAGTCAATATCTTAAAGATCGGAATCTGCGAAAGCGGTGCGCGCAACATCTA 2449375
Qy      377 LysLeuValProValAlaProGlyLysLeuGluGlyArgPheHisGlyProHisValMetSer 396
Db      2449376 CGCATCGCGAC---GATGGCGAGCTGTGTCGCGC-----GTTGGCGTGTATTACAG 2449426
Qy      397 GlyTrpTrpArgAlaProGluGlnAlaAlaGlnAlaPheAspGluGluGlyTrpCys 416
Db      2449427 GGTCTAGCGGCAACGAGCTACACCGAGGCAATTC---ACCAACGCTGTGTTCAAG 2449483
Qy      417 SerGlyAspAlaIleLysLeuAlaAspProAlaAspProGluIleGlyLeuMetCysAsp 436
Db      2449484 ACCGCTGAT-----CTCGGTGCGGTGAGCAAGACCGGCTTTCAGCATCAC 2449531
Qy      437 GlyArgIleAlaGluAspPheLysLeuSerSerGlyValPheValSerValGlyProLeu 456
Db      2449532 GCGCGCAAGAAATTTATTCGTCACCGCGGCGGT-----AAATATGCGCGCGCT 2449585
Qy      457 ArgThrArgAlaValLeuGluGlyGlySerTrpValLeuAspValValAlaAlaPro 476
Db      2449586 GTCTGGAAGACCAAGTGGCGGCCCAACCATGATACAGCGCGGTGTGTTGGGAC 2449645
Qy      477 AspArgGlyCysLeuGlyLeuLeuValPheProArgLeuLeuAspCysArgAlaLeuSer 496
Db      2449646 GCCAAGCCTTCAATCGCGGTATATC-----ACCATGACCTTGAAGCATTCGAG 2449696
Qy      497 Gly-----LeuGlyLysGluAlaSerAsp 504
Db      2449697 GCGTGAAGCAACGACCAAGACAGACAGCTGCGCGTGGCGCATTTGGCCACCGAC 2449756
Qy      505 AlaGluValLeuLysSerGluProValArgAlaTrpPheAlaAspTrpLeuYsArgLeu 524
Db      2449757 CCGGATCTGATTTGCCAGATCGACGCGCC-----GTCAAACAGGCC 2449798
Qy      525 AsnArgGluAlaThrGlyAsnAlaSerArgIleMetTrpValGlyLeuLeuAspTrpPro 544
Db      2449799 AATCTT---CGGTGTCACTAGCGCGAGTGTGCGAAGTTCGAATATCTGCGCGTGCAC 2449855
Qy      545 ProSerIleAspLysGlyGluValThrAspLysGlySerIleAsnGlnArgAlaValLeu 564
Db      2449856 TTCACGAGAGCACCGCGCGAGTACCCGACATGAAAGTCAAAAGAGGTGTGGCC 2449915
Qy      565 GlnTrpArgSerAlaLysValAlaPheAlaLeuTrpArgGlyGlu 578
Db      2449916 GAGAAGTTCCTTCGATATCGAGGCGATCTACACAAAGGAA 2449957

RESULT 15
US-09-396-154-42
; Sequence 42. Application US/09396154
; Patent No. 6602677
; GENERAL INFORMATION:
; APPLICANT: Hall, Mary P.
; TITLE OF INVENTION: Thermostable luciferases and methods of
; FILE REFERENCE: 341.012US1
; CURRENT APPLICATION NUMBER: US/09/396,154
; EARLIER FILING DATE: 1999-09-15
; EARLIER APPLICATION NUMBER: US 09/156, 946
; EARLIER FILING DATE: 1998-09-18

```

QY 89 TyrAlaGlyIleProTyrCysProValSerProAlaTyrSerLeuLeuSerGlnAspLeu 108
|||
Db 2445732 GCTGTGGGT-----GGCGTCAACCGTACCGACTACGAGACCTGTCAGCGAG--- 2445779
|||
QY 109 AlAlaLeuLeuArgHisIleValAlaGlyLeuLeuGlnProGlyLeuValPheAlaIleAspAla 128
|||||
Db 2445780 ---CAGGTGGCTGGTGGTTTACAGAGACTCCGACCGGTGTGTGTTCGCGAACCAC 2445836
|||||
QY 129 AlaProheGlnArgAlaIleGluThrIleLeuProAspAspValProAlaIlePheThr 148
|||
Db 2445837 TCACACGCCGACATAGTCCGCCAA-----CTCTCCGGCAGCGTGGCCGCTCCGGGAG 2445890
|||
QY 149 ArgGlyGluLeuAlaGlyArgThrValSerPheAspSerLeuLeuGlnProGly 168
|||||
Db 2445891 GTACTGCAGATCGCGGTGGGTGGGTCACAGCGCTCATCGGCTCAGCGAGGCGGGCGCC 2445950
|||||
QY 169 GlyIleGluAlaAspAsn-----AlaPheAlaIleThrIleProAsp 182
|||||
Db 2445951 TCAGTGCACCCGCGCGAGCTAACCGCCGCTCCGCTCCGCTACGATCGACGAGCCCGCG 2446010
|||||
QY 183 ThrIleAlaLysPheLeuPheThrSerGlySerThrIleLeuProLysAlaIleProThr 202
|||
Db 2446011 ACGCTT-----ATCTACACTCGGGACACACCGAGACACCAAGGGCTGCCAGTTG 2446061
|||
QY 203 ThrGlnArgMetLeuCysAlaAsnGlnIleMetLeuLeuGlnThrPheProValPheGly 222
|||||
Db 2446062 ACCCAATCCAAACCTGTTACAGATTAAAGGGCGCCAGGGCATATCACCCGACGCTGCTG 2446121
|||||
QY 223 GluGluProProValLeuValAspTrpLeuProTrpAsnHisIlePheGlySerHis 242
|||||
Db 2446122 CGCAAGGGTGAAGCGCGCTGCTGTCTCTGCTCCGCTACGATGTCGTG-----GCCGCG 2446175
|||||
QY 243 AsnIleGlyIleValLeuTyrAsnGlyGlyThrTyrTyrLeuAspArgLysProThr 262
|||
Db 2446176 GCGATCGATGATGGCCGCTTCAC-----TCCAAAGTCAACC 2446211
|||
QY 263 AlaGlnGlyPheAlaGluThrLeuArgAsn-----LeuSerGlnIleSerPro 278
|||||
Db 2446212 GTG---GGATTCAACGACGACATCAAGATCTGCTGCCGATGTTGGGGTTCACACCG 2446268
|||
QY 279 ThrAlaTyrLeuThrValProLysGlyTrpGluGluLeuValGlyAlaLeuGluAsp 298
|||||
Db 2446269 ACCGTGTGTGTGTCGCGCGGAGGTGTTCAGAGAGGTGTACACACCGCGAGCAAGAC 2446328
|||||
QY 299 SerThrLeuArgGluArgPhePheAlaArgMetLysLeuPhePheAlaIle----- 316
|||
Db 2446329 GCCGCCAACGCC-----GCCAAAGGGGGAATCTTCGCGATCGCGCGCAGACC 2446376
|||
QY 317 -----AlaGlyLeu-----Ser 320
|||
Db 2446377 GCGGTGCACTCGAGCGAAGCTTGGACCGCGCGGACCGGGGCTGTACTCGCGGCCAAG 2446436
|||
QY 321 GlnGlyIleTrpAspArgLeu----- 327
|||
Db 2446437 CACGCGGTGTTCACCGCGTGTGTCTACCGCAAGCTGTGCGGACATCGGGTGCACACTGC 2446496
|||
QY 328 -----AspArgValAlaGluGlnHisCysGly 336
|||
Db 2446497 CGGCGCGCTCTCCGCGCGCGCGCGCTGGGTGCGGGGTGTGTCATCTTACCGGGC 2446556
|||
QY 337 GluArgIleArgMetMetAlaGlyLeuGlyMetThrGluThrAlaProSerCysThrPhe 356
|||
Db 2446557 GCGCGTCTCACCATCTACGAGGATATGCTGAGCGGAGCCACATGGGGGCGTCCGCATC 2446616
|||
QY 357 ThrThrGlyProLeuSerMetAlaGlyTyrIleGlyLeuProAlaProGlyCysGluVal 376
|||
Db 2446617 AGCCAGTTCATATCTAAAGATCGAAGCTGGAGAAAGCGGATGCCCGGCAACAAGCTTA 2446676
|||
QY 377 LysLeuValProValAspGlyLysLeuGluGlyArgPheHisGlyProHisValMetSer 396
|||
Db 2446677 CGCATGCCCGAC---GATGCGGAGGTGCTGTGTGCG---GGTGGCGTGTGATTTCAGC 2446727

QY 397 GlyTyrTrpArgAlaProGluGlnAsnAlaGlnAlaPheAspGluGluGlyTyrTyrCys 416
|||
Db 2446728 GGCTACTGGCGCAACGAGCGGTATCACCGAGCATTC---ACCGACGGCTGTGTCAG 2446784
|||
QY 417 SerGlyAspAlaIleLysLeuAlaAspProAlaAspProGlnIleGlyLeuMetPheAsp 436
|||
Db 2446785 ACCGGTAT-----CTCGGTGCGGTGACCAAGACGGGTTCTTGACATCAC 2446832
|||
QY 437 GlyArgIleAlaGluAspPheLysLeuSerSerGlyValPheValSerValGlyProLeu 456
|||
Db 2446833 GGCCGCAAGAAAGAAATATATCGTACACCGCGGGGT-----AAAATGTCCGCCCGCT 2446886
|||
QY 457 ArgThrArgAlaValLeuGluGlySerTyrValLeuAspValValAlaIlePro 476
|||
Db 2446887 GTCTGGAGAACCAAGCTCGCGGCCACCACTGATCAGCCAGCGGTGTGTGGTGGGAC 2446946
|||
QY 477 AspArgGluCysLeuGlyLeuLeuValPheProArgLeuLeuAspCysArgAlaLeuSer 496
|||
Db 2446947 GCCAAGCCTTCATCGCGGCTTAATC-----ACCATTCAGCCTTAGAGATTGAG 2446997
|||
QY 497 Gly-----LeuGlyLysGluAlaSerAsp 504
|||
Db 2446998 GGCTGAGAGCAACGACAGCAAGACAGTGGCGCGTGGGTGGCATTTGGCCACCGAC 2447057
|||
QY 505 AlaGluValLeuAlaSerGluProValArgAlaIleTrpPheAlaAspTrpLeuLysArgLeu 524
|||||
Db 2447058 CCGATCTGATGCTCCGATCGACGCGGCC-----GTCAAAACAGGCC 2447099
|||||
QY 525 AsnArgGluAlaThrGlyAsnAlaSerArgIleMetTrpValGlyLeuLeuAspTrpPro 544
|||||
Db 2447100 AATCTT---GCGGTGTCAATGCTCGAGTGCAGAGTTCCGAAATATCCGCTGCAC 2447156
|||||
QY 545 ProSerIleAspLysGlyGluValThrAspLysGlySerIleAsnGlnArgAlaValLeu 564
|||
Db 2447157 TTACCGAGAGACACCGCGAGCTGACCCGACATGAAGTCAACCAAGGTGTGTGCC 2447216
|||
QY 565 GlnTrpArgSerAlaLysValAspAlaLeuTyrArgGlyGlu 578
|||
Db 2447217 GAGAAAGTTCCTTCGATATCGAGGCGATCTCAACAAAGAA 2447258
|||
RESULT 14
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37RV
US-09-103-840A-1
Alignment Scores:
Pred. No.: 3,68e-13 Length: 4411529
Score: 259.00 Matches: 142
Percent Similarity: 37.92% Conservative: 106
Best Local Similarity: 21.71% Mismatches: 282
Query Match: 8.40% Indels: 124
DB: Gaps: 26
US-09-750-986d-30 (1-589) x US-09-103-840A-1 (1-4411529)

```

Db      425 TCRAATTAATACTATGAACTATATATATAGACTTAATGAAGCTTAGAGGCT 484
Qy      170 TLeGuaLaasphenaLaPhaLaalathrGlyProaspThr----- 183
Db      485 TATCAATGCTCAACAACCTTATTTTCATAAATTCCTAGTAGTATCGAGTAAAAA 544
Qy      184 -----TLeLaLysPhaLeuPhethrSergly 192
Db      545 TTTAAACATATCTTTTAAATCGAGACGATCGAGTTGATGATTTCTTCTGCT 604
Qy      193 SerThrLysLeuProLysalaValProThrThrGlnaMetLeuCyalaamGlnGln 212
Db      605 ACAACGTGCTGCGGAAGGAGTCACTAATCAACAAGAAATATGTGCA-----CGA 658
Qy      213 MetLeuLeuGlnThrPheProvalPheGly-----GluGluProProvalLeuVal 229
Db      659 TTTTCTCTTGCAAAAGATCTTACTTGTGTAACGCAATTATCCACGACGCAATTTTA 718
Qy      230 AspThrLeuProThrPasnHisThrPheGlyGlySerHisasnIleGly----- 245
Db      719 ACCGTAATACCTTCCACCATGTTTGTATGATGACACCATTAAGATTACTTACTTGT 778
Qy      246 -----TLeValLeuTyraasnGlyGlyThrTyrrLeuaspaspGlyLysProThr 262
Db      779 CGATTCGAGTTGTTCTAATGCAC----- 802
Qy      263 AlaGlnGlyPhealagluThr-----LeuargPasnLeuSerglyLysProThrAla 280
Db      803 -----ACGTTGAGAAAACTATTCTCAATCATTAACAGATTATAAGTGAAGT 856
Qy      281 TyrrLeuThrValProLysGlyTrpGluGluLeuValGlyLalaLeuGuaaspSerThr 300
Db      857 ACTTTACTGTGACCA-----ACATTAATGCGCATTTCTTGCAAAAAAGTCA 904
Qy      301 LeuargGluargPhePhealAargMetLysLeuPhePhePhealAalaalaglyLeuSer 320
Db      905 GTTGAAAGATGACATTATGACACTTAAGAAATTCATCGTGTGCGCACCTTATCA 964
Qy      321 GlnGlyTlEtrpaspargLeuaspargValAlaGluGlnHisCybglyGluValGlyLeuArg 340
Db      965 AAGAAAT-----GGGGAGATGCTGAAA 988
Qy      341 -----MetMetAlaGlyLeuGlyMetThrGluThrAlaProSer 353
Db      989 AAACGGTTTAATTAACCTTGTGACGCAAGGATGATTAACAGAAACCACTCGGCT 1048
Qy      354 CyethrPheThrThrGlyProLeuSerMetAlaGlyTyrrIleGlyLeuProalProGly 373
Db      1049 GTTTTATTATCACCGAAAGGTGACGCCAAACCGGATCACTGGTAAATAGTACATTT 1108
Qy      374 CyagGluValLysLeuVal---ProvalaspGlyLysLeuGluGly----- 387
Db      1109 CAGCGTGTAAAGTTGTCGATCCTACAAACAGAAAAATTTTGGGCCCAATGAACCTGA 1168
Qy      388 -----ArgPheHisGlyProHisValMetSerglyTyrrTrpArgAlaProGluGlnasn 405
Db      1169 GAATTGTATTTTAAAGCCCGATGATATGAAGGTTATTATAATGAAGAACTACT 1228
Qy      406 AlaGlnAlaPheaspGluGluGlyTyrrCyaserglyAspAlaIleLysLeuAlaasp 425
Db      1229 AAGGCAATTATGTATATGACGAGTGTGCGCTGTGAT---ATTGCTTATTATGAC 1285
Qy      426 ProAlaaspProGlnLysGlyLeuMetPheaspGlyValGllealagluaspPheLysLeu 445
Db      1286 AAT-----GATGGCCATTTTATATATGTCAGCGCTG 1318
Qy      446 SerSerglyVal-----PheValSerValGlyProLeuargThrArgAlaValIleu 462
Db      1319 AAGTCACTGATTAAATTAAGGTTATCAGGTTGCACCTGCTGAATTTGAGGAATATCTC 1378
Qy      463 GluGlyGlySerTyrrValLeuaspValValValAlaAlaPheaspargGluCybLeuGly 482

```

```

Db      1379 TTACAATCCGATATTTGTTGATCCGCGGCTTACTGTATACCGGATGACCGCGGCG 1438
Qy      483 LeuLeuValPheProargLeuLeuaspCyargAlaLeuSerglyLeuGlyLysGluAla 502
Db      1439 -----GAGCTTCC-ACGTCAAGGTGTGTAGT----- 1464
Qy      503 SeraspAlaGluValLeuAlaSerGluProvalArgAlaThrPheAlaaspTrp----- 520
Db      1465 -----ACAGACTGAAAAATA 1479
Qy      521 LeuLysargLeuasnarg-----GluAlaThrGlyAsnAlaSerargIle 535
Db      1480 TCRAACGAACAATGTAACAGATTATGTGCGCATCAAGTTCAACAGCCAAATGCGCT 1539
Qy      536 MetTrpValGly 539
Db      1540 ACGTGTGGGCT 1551

RESULT 13
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Alignment Scores:
Pred. No.: 3,676-13 Length: 4403765
Score: 259.00 Matches: 142
Percent Similarity: 37.92% Conservative: 106
Best Local Similarity: 21.71% Mismatches: 282
Query Match: 8.40% Indels: 124
DB: 3 Gaps: 26

US-09-750-986D-30 (1-589) x US-09-103-840A-2 (1-4403765)
Qy      2 ArgSerLeuGluValLeuLeu-----ProPheProGlyArgIleLeu 15
Db      2445438 AGGTCAAGTAGAGGTGTGATCGGCGGTGAATTAAGGCTCCCGCCCATTACTACGTGCGC 2445497
Qy      16 GluArg-----LeuGlnHisTrpAlaLysThrArgProGluGln 28
Db      2445498 GAGCAGACGACCAACCTCCGCGCATGCTGTTTCGACAT---GAACGTACAGATCCCGACATAC 2445554
Qy      29 ThrCyValAlaAlaArgAlaAlaLeuGluGlyTrpArgArgIleSerTyrrAlaGluMet 40
Db      2445555 GTC---ATCATACAGCTGATGACGCGGCTGTGACCGATGTCACGTTGCGAGGCA 2445611
Qy      49 PheHisAsnValArgAlaIlealaglnSerLeuLeuProTyrrGlyLeuSerAlaGluArg 68
Db      2445612 GCCAACCAAGTTGTGCTCCGCGGCTTCGCGTTTATTACTGCGGCGGTGACAGCGCGCAT 2445671
Qy      69 ProLeuLeuIleValSerglyAsnaspLeuGluHisLeuGlnLeuAlaPheGlyAlaMet 88
Db      2445672 CGGTAAGTCACTTCTTCTGCAACCCGCTACGAGTGGCGCATCTCGATTTCGCGGATTCTG 2445731

```

Db 545 TTAAACCTATTCTTTAATCGACGATCGTGGCATTAATGTTCTTCGTGT 604
Qy 193 SerThyLeuProLysAlaValProThrThrGlnMetLeuCysAlaAsnGln 212
Db 605 ACAACTGGTGTGCGGAGGAGTCACTGTAACCAAGAAATATGTTGCA-----CGA 658
Qy 213 MetLeuGlnThrPheProValPheGly-----GlnGluProProValLeuVal 229
Db 659 TTTCTATTGCAAAAGATCTTACTTTTGTAAAGCAATTAATCCACGTCAGCAATTTTA 718
Qy 230 AspThrLeuProThrPheAsnHisThrPheGlyGlySerHisAsnIleGly----- 245
Db 719 ACGGTAAATACCTTTCCACCATGATGTTGTATGATGACACCATTAAGATTAATCTTGT 778
Qy 246 -----IleValLeuTyraAsnGlyGlyThrTyrtTyrtLeuAspArgLysPheProThr 262
Db 779 GGATTCGAGTGTGTTCTTAATGCAC----- 802
Qy 263 AlaGlnGlyPheAlaGlnThr-----LeuArgAsnLeuSerGluIleSerProThrAla 280
Db 803 -----ACGTTGAAGAAAACTATTCTTCAATCAATTAAGATTAATTAAGTGAAGT 856
Qy 281 TyrtLeuThrValProLysGlyTyrGlnGluLeuValGlyAlaLeuGluAspSerThr 300
Db 857 ACTTACTTGTACCA-----ACATTAATGCGCATTTCTTGCAAAAAGTGCATTA 904
Qy 301 LeuArgGluArgPhePheAlaArgMetLysLeuPhePheAlaAlaIleGlyLeuSer 320
Db 905 GTTAAAAAGTACGATTAATGCGACTTAAGAAATTCGATCGTGGCGACCTTATCA 964
Qy 321 GlnGlyIleTyrAspArgLeuAspArgValAlaGlnGlnHisCysGlyGluArgIleArg 340
Db 965 AAGAAAT-----GGGAGATGTGTAA 988
Qy 341 -----MetMetAlaGlyLeuGlyMetThrGluThrAlaProSer 353
Db 989 AAACGGTTTAATTAACCTTTGTCAGGCAAGGATGATTAACAGAAACACTTCGCGCT 1048
Qy 354 CysThrPheThrThrGlyProLeuSerMetAlaGlyTyrtIleGlyLeuProAlaProGly 373
Db 1049 GTTTTAATTAACCGAAAGGTGACGCCAAACCGGATCAACTGTTAAATTAATGATCACTTA 1108
Qy 374 CysGlnValLysLeuVal-----ProValAspGlyLysLeuGlnGly----- 387
Db 1109 CAGCGCTTAAAGTGTGCATCTTCAACACGAAAAATTTGGGCGCAATGAACTGGA 1168
Qy 388 -----ArgPheHisGlyProHisValMetSerGlyTyrtTyrArgAlaProGlnGlnAsn 405
Db 1169 GAATGATTTTAAAGCCGATGATGATGAAGGTTATTAATTAATGAAGAACTACT 1228
Qy 406 AlaGlnAlaPheAspGlnGlnGlyTyrtTyrtCysSerGlyAspAlaIleLysLeuAlaAsp 425
Db 1229 AAAGCAATTAATTAATGAAGTGTGCGCTCTGTATAT--ATTGCTTATTAATGAC 1285
Qy 426 ProAlaAspProGlnLysGlyLeuMetPheAspGlyArgIleAlaGluAspPheLysLeu 445
Db 1286 AAT-----GATGCGCATTTTATTAATATGTCGACAGGCTG 1318
Qy 446 SerSerGlyVal-----PheValSerValGlyProLeuArgThrArgAlaValLeu 462
Db 1319 AAGTCACTGATTAATTAAGTTATCAAGTTGTCACCTGCTGAATTAAGGGAATATCTC 1378
Qy 463 GlnGlyGlySerTyrtValLeuAspValValAlaAlaProAspArgGlnCysLeuGly 482
Db 1379 TTACAAATCCGATATGTTGATGCGCGGTTACTGTTATCCGATGAAGCGCGGCG 1438
Qy 483 LeuLeuValPheProArgLeuAspCysArgAlaLeuSerGlyLeuGlyLysGluAla 502
Db 1439 -----GAGCTTCC-AGCTGCAAGGTGTTGAT----- 1464
Qy 503 SerAspAlaGluValLeuAlaSerGluProValArgAlaIleThrPheAlaAspTyr----- 520

Db 1465 -----ACAGACTGGAATA 1479
Qy 521 LeuLysArgLeuAsnArg-----GlnAlaThrGlyAsnAlaSerArgIle 535
Db 1480 TCTAAAGCAAAATCTTACAAAGATTAATGTTGCGCAGTCAAGTTTCAACAGCAAAATGCT 1539
Qy 536 MetTyrValGly 539
Db 1540 ACGTGTGGGGT 1551

RESULT 12
US-09-396-154-11
Sequence 11, Application US/09396154
Patent No. 6602677
GENERAL INFORMATION:
APPLICANT: Wood, Keith V.
TITLE OF INVENTION: Thermostable luciferases and methods of
FILE REFERENCE: 341.012US1
CURRENT APPLICATION NUMBER: US/09/396,154
EARLIER FILING DATE: 1999-09-15
EARLIER APPLICATION NUMBER: US 09/156,946
EARLIER FILING DATE: 1998-09-18
EARLIER APPLICATION NUMBER: PCT/US98/19494
EARLIER FILING DATE: 1998-09-18
EARLIER APPLICATION NUMBER: US 60/059,379
NUMBER OF SEQ ID NOS: 93
SOFTWARE: FastsEQ for Windows Version 3.0
SEQ ID NO 11
LENGTH: 1639
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Mutant luciferase
US-09-396-154-11

Alignment Scores:
Pred. No.: 3,37e-19 Length: 1639
Score: 262.50 Matches: 129
Percent Similarity: 36.52% Conservative: 77
Best Local Similarity: 22.87% Mismatches: 218
Query Match: 8.52% Indels: 141
DB: 4 Gaps: 23

US-09-750-986D-30 (1-589) x US-09-396-154-11 (1-1639)
Qy 30 CysValAlaAlaIleArgAlaIleAsnGlyGluTyrArgArgIleSerTyrtAlaGlnMetPhe 49
Db 119 TGCATAGCATTTGCAAAATGCTCATACAAA--GAAAATGTTTATATGAAGATTTCG 175
Qy 50 HisAsnValArgAlaIleAlaGlnSerLeuProTyrtGlyLeuSerAlaGluArgPro 69
Db 176 AAACGTCGTGTGTTTACGCGGAAAGTTTAAAGATGATGATTAACAAACAGACACA 235
Qy 70 LeuLeuIleValSerGlyAsnAspLeuGlnHisLeuGlnLeuAlaPheGlyAlaMetTyrt 89
Db 236 ATAGCGGTGTGAGCAAAATGCTGCAATTTTCTTCCTGTAATTCGATCATTTGTAT 295
Qy 90 AlaGlyIleProTyrtCysProValSerProAlaTyrtSerLeuLeuSerGlnAspLeuAla 109
Db 296 CTGGAAATATTTGTGGCACCCTGTTAAACGATTAATAC-----ATTGAACGTGAATTAATA 349
Qy 110 LysLeuArgHisIleValGlyLeuLeuGlnProGlyLeuValPheAlaIleAspAlaIle 129
Db 350 -----CACAGCTCTGGTATGTAAACACGCGATGTTTGTTCCTCCAGAAATACT 400
Qy 130 ProPheGlnArgAlaIleGluThrIleLeuProAspAspValProAlaIlePheThrArg 149
Db 401 -----TTTCAAAAATGCTGAAT-----GTAATA 424
Qy 150 GlyGluLeuAlaIleTyrtArgThrValSerPheAspSerLeuLeuGlnInProGlyGly 169

```

Db      659 TTTTCTTGCAAAAGATCTTACTTTGGTAACGCAATTATCCACGACGACATTTTA 718
Qy      230 AepTrpLeuProTrpAsnHisThrPheGlySerHisAsnIleGly----- 245
Db      719 ACGGTAACTCTTCCACCATGTTTGGTATGATGACCACTTAGACTTACTTGT 778
Qy      246 -----1LeuValLeuTyrAsnGlyGlyThrTyrTrpLeuAspArgLysProThr 262
Db      779 GGATCCGAGTGTGTCTAATGAC----- 802
Qy      263 AlaGlnGlyPheAlaGluThr-----LeuArgAsnLeuSerGluIleSerProThrAla 280
Db      803 -----ACGTTTGAAGAAAACTATTTCTACATCATTAACAAGTTTAAAGTGAAGT 856
Qy      281 TyrLeuThrValProLysGlyTyrGlnGluLeuValGlyAlaLeuGlnArgAspSerThr 300
Db      857 ACTTTACTGTACCA-----ACATTTAATGCACTTTCTTGCAAAAAGTGCATTA 904
Qy      301 LeuArgGlnArgPhePheAlaArgMetLysLeuPhePheHeAlaAlaAlaGlyLeuSer 320
Db      905 GTTGAAAAGTACATTTATGACATTAAAGAAATTGCACTGCGCGCACTTATCA 964
Qy      321 GlnGlyIleTrpAspArgLeuAspArgValAlaGlnGlnHisCysGlyGluArgIleArg 340
Db      965 AAAGAATT-----GGCGAGATGCTGAAA 988
Qy      341 -----MetMetAlaGlyLeuGlyMetThrGluThrAlaProSer 353
Db      989 AAACGGTTTAAATTAACTTTGTCAGGCAAGGATGATTAACAGAAACCACTTCGGCT 1048
Qy      354 CysThrPheThrThrGlyProLeuSerMetAlaGlyTyrIleGlyLeuProAlaProGly 373
Db      1049 GTTTTAATTACCCGAAANNNNNNCCAGACCGGATCAACCTGTAATAATGATCACTTT 1108
Qy      374 CysGlnValLysLeuVal-----ProValAspGlyLysLeuGlnGly----- 387
Db      1109 CAGCGTTTAAAGTGTGCATCTCACAACAGAAAAATTTTGGCGCAAAATGAACCTGGA 1168
Qy      388 -----ArgPheHisGlyProHisValMetSerGlyTyrTrpArgAlaProGlnGlnAsn 405
Db      1169 GAATTGATTATTAAGCCCGCATGATATGAGGGTTATTAATTAATGAAGACTACT 1228
Qy      406 AlaGlnAlaPheAspGlnGlnGlyTyrTyrCysSerGlyAspAlaIleLysLeuAlaAsp 425
Db      1229 AAAGCAATTATTAATGAACGATGCGCTGCTGCTGATATT----- 1273
Qy      426 ProAlaAspProGlnLysGlyLeuMetPheAspGlyArgIleAlaGluAspPheLysLeu 445
Db      1274 -----GCTTATTATGAC----- 1285
Qy      446 SerSerGlyValPheValSerValGlyProLeuArgThrArgAlaValLeuGlnGlyGly 465
Db      1286 AATGATGGCATTATTTATTTATGTCGACAGCGCTGAAGTCATTAAATTAATGAAGT--- 1342
Qy      466 SerTyrValLeuAspValValAlaAlaProAspArgGlnCysLeuGlyLeuLeuVal 485
Db      1343 -----TATCAGGTTGCACCTGCTGAATTAAGGGAATA---CTCTTACAA 1384
Qy      486 PheProArgLeuLeuAspCysArgAlaLeuSerGlyLeuGlyLysGlnAlaSerAspAla 505
Db      1385 CATCCGTAATTTGTGATGCG---GGCGTTACTGTTATCCGATGAAGCCGCGCGAG 1441
Qy      506 GluValLeuAlaSerGlnProValArgAlaTrpPheAlaAspTrpLeuLysArgLeuAsn 525
Db      1442 CTTCACAGCTCAGGTGTGTGATGACAGACTGA-----AATATCTTAAC 1486
Qy      526 ArgGlnAlaThrGlyAsn-----AlaSerArgIleMetCys--- 537
Db      1487 GAACAATAATCGTACAAATTTTGTTCAGTCACATTTCAACAGCAAAATGCTACGTTGT 1546
Qy      538 ---ValGlyLeuLeuAspTrpProProSerIleAspLysGlyValAlaThrAspLys 555

```

```

Db      1547 GGGGTGAATTTTGGATGAATTCACCAAGATCACTGAAAAAATTGACAGAAAA 1603
RESULT 11
US-09-396-154-43
/ Sequence 43, Application US/09396154
/ Patent No. 6602677
/ GENERAL INFORMATION:
/ APPLICANT: Wood, Keith V.
/ APPLICANT: Hall, Mary P.
/ TITLE OF INVENTION: Thermostable luciferases and methods of
/ TITLE OF INVENTION: Production
/ FILE REFERENCE: 341.012US1
/ CURRENT APPLICATION NUMBER: US/09/396,154
/ EARLIER FILING DATE: 1999-09-15
/ EARLIER APPLICATION NUMBER: US 09/156,946
/ EARLIER FILING DATE: 1998-09-18
/ EARLIER APPLICATION NUMBER: PCT/US98/19494
/ EARLIER FILING DATE: 1998-09-18
/ EARLIER APPLICATION NUMBER: US 60/059,379
/ EARLIER FILING DATE: 1997-09-19
/ NUMBER OF SEQ. ID NOS: 93
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ. ID NO. 43
/ LENGTH: 1639
/ TYPE: DNA
/ ORGANISM: Artificial Sequence
/ OTHER INFORMATION: Mutant luciferase
US-09-396-154-43

Alignment Scores:
Pred. No.: 2,61e-19 Length: 1639
Score: 263.50 Matches: 128
Percent Similarity: 36.52% Conservative: 78
Best Local Similarity: 22.70% Mismatches: 218
Query Match: 8.55% Indels: 141
DB: Gaps: 23

US-09-750-986d-30 (1-569) x US-09-396-154-43 (1-1639)
Qy      30 CysValAlaAlaArgAlaAlaAsnGlyGluTrpArgArgIleSerTyrAlaGluMetCys 49
Db      119 TGCATGACATTTGACAAATGCTCTCATACAAA---GAAAACTTTTATATGACAGGTTTCTG 175
Qy      50 HisAsnValArgAlaAlaIleAlaGlnSerLeuLeuProTyrGlyLeuSerAlaGluArgPro 69
Db      176 AAACGTGCTGTCGTTTACCGGAAAGTTTAAAGATGATGATTAACAAAACGACACA 235
Qy      70 LeuLeuIleValSerGlyAsnAspLeuGlnHisLeuGlnLeuAlaPheGlyAlaMetCys 89
Db      236 ATAGCGGTGTGTCGCAAAATAGCTGCAATTTTCTTCCTGTAATGATGATGATGAT 295
Qy      90 AlaGlyIleProTyrCysProValSerProAlaTyrSerLeuLeuSerGlnAspLeuAla 109
Db      296 CTGGAATATATTTGGACACTGTTAAGATTAATAC-----ATTGAACGTGAATTAATA 349
Qy      110 LysLeuArgHisIleValGlyLeuLeuGlnProGlyLeuValPheAlaAlaAspAlaAla 129
Db      350 -----CACGTGTGTGATTTGTAACCCGCAATGATGTTTTCCTCCAAAGAACTACT 400
Qy      130 PheGlnAlaArgAlaIleGluThrIleLeuProAspAspValProAlaIlePheThrArg 149
Db      401 -----TTTCAAAAATACTGAAAT-----GTAAAA 424
Qy      150 GlyGlnLeuAlaGlyArgThrValSerPheAspSerLeuGlnGlnProGlyGly 169
Db      425 TCTAAATTAATAATCTATTGTAACCTATTAATTAATTAATTAATGAAGACTGAGAGT 484
Qy      170 IleGlnAlaAspAsnAlaPheAlaIleThrGlyProAspThr----- 183
Db      485 TATCAATGCTCAACAACATTATTTCTCAAAATTCGATAGTAATGACGTAACGTAACAAA 544
Qy      184 -----IleAlaLysPheLeuPheThrSerGly 192

```

QY 319 LeuSerGlnGlyIle-----TTP-----AspArgLeuAspArg 329
Db 991 CGTTCAATTAGCGCGTATCTCAATGATGTTATCGCTTTCGCGATAAAGATCTTCAA 1050
QY 330 ValAlaGluGlnHisCysGlyValArgIleArgMetMetAla----- 343
Db 1051 CCACATACGTAAATTTTAGGTGGCGAGTTCGTTTTCCTGCGCGAGCGCTCGTTTA 1110
QY 344 -----GlyLeuGly 346
Db 1111 GATGATCGCGTATCGCGCTTTTCTTATGCTGCTATTAATAATAATTAATGTTATGCG 1170
QY 347 MetThrGluThrAlaProSerCysThrPheThrThrGlyProLeuSerMetAlaGlyTyr 366
Db 1171 ATGACAGAGACTTGTGCGACGATCTTGTGGAGAAACCATTTATTAAGTACGCTCT 1230
QY 367 IleGlyLeuProAlaProGlyCysGlyValLysLeuValProValAspGlyLysLeuGlu 366
Db 1231 ATTTGTACTCCGTTACCGGGGGTTGAGTACGT-----ATAGCGAGAAAGAAATGAG 1281
QY 387 GlyArgPheHisGlyProHisValMetSerGlyTyrTrpArgAlaProGluGlnAsnAla 406
Db 1282 ATCCAGTGGCGCGTACGTATCGTATGAAAGCTTATTTAATAAGCTGAGAGACAGCTT 1341
QY 407 GlnAlaPheAspGluGluGlyTyrTyrCysSerGlyAspAlaIleLysLeuAlaAspPro 426
Db 1342 GCTGCGTTTACTGAAAGATGCTGTGTTACGTACCGGTATCGGGGTGATTA----- 1392
QY 427 AlaAspProGlnLysGlyLeuMetPheAspGlyArgIleAlaGluAspPheLysSer 446
Db 1393 --GATGAGATGGGATGTTATATTATTAACGAACGTTTGAAGATCTTATGAAACCTCA 1449
QY 447 SerGlyValPheValSerValGlyProLeuArgThrArgAlaValLeuGluGlyLysSer 466
Db 1450 AATGGAATAATATATGCA-----CCGCAATGATTTGAAGTACATTAGGCGAAGATCGC 1503
QY 467 TyrValLeuAspValValValAlaAlaProAspArgGluCysLeuGlyLeuValPhe 486
Db 1504 TTATATGAGCATATCGCTGTGATTTGACATATCCGTAATATTTGTTCTGCACTGATGTA 1563
QY 487 ProArgLeuLeuAspCys-----ArgAlaLeuSerGlyLeu 498
Db 1564 CCT-----TGCCTTGTATGCAATTAAGAAACATGCTCGAGCATTTAAT----- 1605
QY 499 GlyLysGluAlaSerAspAlaGluValLeuAlaSerGluProValArgAlaTrpPheAla 518
Db 1606 TTTAAATATATACATCGGATAGATTAATTAAGCCATATCTAAGATCAAGAAATTTTGTAT 1665
QY 519 AspTrpLeuLysArgLeuAsnArgGluAlaThrGlyAsnAlaSerArgIleMetTrpVal 538
Db 1666 GAACGCTTGTGTAATGCAACGTAAT---TTGTCTAGTTTCATCAAGTAAACGATTT 1722
QY 539 GlyLeuLeuAspThrProProSerIleAspLysGlyGluValThrAspLysGlySerIle 558
Db 1723 ACCCTATTAACGGAAGGTTTACGATGAGTCTGAGAAATTTGACA----- 1767
QY 559 AsnGlnArgAlaValLeuGlnTrpArgSerAlaLysValAspAlaLeuTyrArgGlyGlu 578
Db 1768 -----CTTACCTTTAAATTAAGACGTAAATAATTTCTGACGCTATCGAAATGAA 1818
QY 579 AspGlnSerMetLeuArgAsp 585
Db 1819 ATGAGACAGATGTATCAAGAG 1839

RESULT 10
US-09-396-154-10
; Sequence 10, Application US/09396154
; Patent No. 6602677
; GENERAL INFORMATION:
; APPLICANT: Wood, Keith V.
; APPLICANT: Hall, Mary P.
; TITLE OF INVENTION: Thermostable luciferases and methods of
; TITLE OF INVENTION: production

FILE REFERENCE: 341.012US1
CURRENT APPLICATION NUMBER: US/09/396,154
CURRENT FILING DATE: 1999-09-15
EARLIER APPLICATION NUMBER: US 09/156,946
EARLIER FILING DATE: 1998-09-18
EARLIER APPLICATION NUMBER: PCT/US98/19494
EARLIER FILING DATE: 1998-09-18
EARLIER APPLICATION NUMBER: US 60/059,379
EARLIER FILING DATE: 1997-09-19
NUMBER OF SEQ ID NOS: 93
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1039
LENGTH: 1639
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
NAME/KEY: unsure
LOCATION: (1067)..(1072)
OTHER INFORMATION: Unknown nucleotides
OTHER INFORMATION: Mutant luciferase
US-09-396-154-10
Alignment Scores:
Pred. No.: 2,61e-19 Length: 1639
Score: 263.50 Matches: 132
Percent Similarity: 37.31% Conservative: 84
Best Local Similarity: 22.80% Mismatches: 226
Query Match: 8.55% Indels: 137
DB: Gaps: 24
US-09-750-986D-30 (1-589) x US-09-396-154-10 (1-1639)
QY 30 CysValAlaAlaArgAlaAlaAsnGlyGluTrpArgArgIleSerTyrAlaGluMetPhe 49
Db 119 TGCATATGCAATGCAATGCTCATACAAA---GCCCGTGTATATATGAGAGTTTTA 175
QY 50 HisAsnValArgAlaIleAlaGlnSerLeuLeuProTyrGlyLeuSerAlaGluArgPro 69
Db 176 AATGTCTGTCGTGTAGGCGGAAGTTTAAAGATATGATTAAGTAAACAAACGACACA 235
QY 70 LeuLeuIleValSerGlyAsnAspLeuGlnHisLeuGlnLeuAlaPheGlyAlaMetTyr 89
Db 236 ATAGCGGTGTGAGCAAAATGGTTGCAATTTTCTTCCTGTAATTCGATCATTTGTAT 295
QY 90 AlaGlyIleProTyrCysProValSerProAlaTyrSerLeuLeuSerGlnAspLeuAla 109
Db 296 CTGGAAATTAATTTGGCACTGTTAAACGATAATAC-----ATGACCTGAAATTAATA 349
QY 110 LysLeuArgHisIleValGlyLeuLeuGlnProGlyLeuValPheAlaAspAlaAla 129
Db 350 -----CACAGTCTTGGTATTTGTAACCCAGCATAGTTTTTGTCCAGAAATACT 400
QY 130 ProPheGlnArgAlaIleGluThrIleLeuProAspAspValProAlaIlePheThrArg 149
Db 401 ---TTTCAAAAGTACTGAAT-----GTAATA 424
QY 150 GlyLysLeuAlaGlyArgArgThrValSerPheAspSerLeuLeuGlnProGlyGly 169
Db 425 TCTAAATTAATCTGTAAGAACTATTAATTAATTAAGTAAATGAAGACTTAAGAGGT 484
QY 170 IleGluAlaAspAsnAlaPheAlaAlaArgGly----- 180
Db 485 TATCAATGCTCTCAACAACCTTATTTCTCAAAATTCGATTAATTCGACGTAATAAAAA 544
QY 181 -----ProAspThrIleAlaLysPheLeuPheThrSerGly 192
Db 545 TTTAAACATATTTCTTTATTCGAGACGATACAGGTTGCGTTATGTTCTTCTGTGT 604
QY 193 SerThrLysLeuProLysAlaValProThrThrGlnArgMetLeuCysAlaAsnGlnGln 212
Db 605 ACAACTGCTGTCGGAAGGAGTCATGCTAATCTACACAAAGATATTTGTGCA-----CGA 658
QY 213 MetLeuLeuGlnThrPheProValPheGly-----GluGluProProValLeuVal 229

```

QY 355 -----ThrpheThrThrglyProLeuSerMetAlaGlyTyrIleGlyLeuProAlaPro 372
DB 1774 TGGCATGATTTCCAAATTAACCCAAATTC-----ATCGCACACATGATGCCA 1821
QY 373 GLYCAGLValIlyLeuValProValAspGlyLysLeuGluGlyArgPheHISGlyPro 392
DB 1822 AAGCGGAATGAAA-----ATTGGGGAATAATGAAATCCTTGTGCGTGGCGGA 1872
QY 393 HlValMetSerGlyTyrTrpArgAlaProGluGlnAsnAlaGlnAlaPheAspGluGlu 412
DB 1873 ATGCTGATGAAGGCTTATCAAGAGCCAGAGAAAGCGCTCAAGCTTCAACGAAGAT 1932
QY 413 GLYTrpTyrCysSerGlyAspAlaIleLysLeuAlaAspProAlaAspProGlnLysGly 432
DB 1933 GGTTTCTTAAAACTGGCGATGCGAGAAATTT-----GACGAACAAGCGAAT 1980
QY 433 LeuMetPheAspGlyArgIleAlaGluAspPheLysLeuSerSerGlyValPheValSer 452
DB 1981 TTAATTATTAATCCATCCTATCAAGAAATTAATGAAACCTCAACGCAATATATCGCA 2040
QY 453 ValGlyProLeuArgThrArgAlaValLeuGluGlyLysSerTyrValLeuAspValVal 472
DB 2041 -----CCACAAATATATCGAAAGCAAAATCGTAAGATTAATTTATCGAAATTCGCG 2094
QY 473 ValAlaAlaProAspArgGlyCysLeuGlyLeuLeuValPheProArgLeuLeuAspCys 492
DB 2095 ATCATGCTGATCGRAAAATATGATTCGCGCTTATTTGCT-----TGC 2142
QY 493 ---ArgAlaLeuSerGlyLeuGlyLysGluAlaSer-----AspAlaGlu 506
DB 2143 TTGTATATGTTTGAAGAAATACGCTAAACAGCTCAATTTAAATATCATGACCGTTAGAA 2202
QY 507 ValLeuAlaSerGluProValArgAlaIleTrpPheAlaAspTrpLeuLysArgLeuAsnArg 526
DB 2203 CTACTTAAAAATCTGACATTCGAAATGTTGAGCATGCTATTAATGCGGTGCAAAA 2262
QY 527 GluAlaThrGlyAsnAlaSerArgIleMetTrpValGlyLeuLeuAspTrpPropioSer 546
DB 2263 GAA---TTGGCTCATTTTCGAGCAAGTAAATAATTCACGTTACTTCTCAAGCATTCAGC 2319
QY 547 IleAspLysGlyGluValThrAspLysGlySerIleAsnGlnArgAlaValLeuGlnTrp 566
DB 2320 ATTAAATTAGCGCAAAATTACACCAACTTAAATTAAGTAAAGATTTTGAACGT 2379
QY 567 ArgSerAlaLysValAspAlaLeuTyrArgGlyGluAsp 579
DB 2380 TATCGACGAAATGAGCAATGATTCATTCACAAAGA 2418

RESULT 9
US-09-543-681A-2894
; Sequence 2894, Application US/09543681A
; Patent No. 6605708
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543, 681A
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 2894
; LENGTH: 1845
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-2894

Alignment Scores:
Score: 1.66e-19 Length: 1845
Percent Similarity: 26.00 Matches: 140
Beet Local Similarity: 37.87% Conservative: 105
Mismatch: 21.64% Mismatches: 270

```

```

Query Match: 8.63% Indels: 132
DB: 4 Gaps: 23
US-09-750-986D-30 (1-589) x US-09-543-681A-2894 (1-1845)

QY 14 IleLeuGluArgLeuGlnHisTrpAlaLysThrArgProGluGlnThrCysValAlaAla 33
DB 70 ATAATTAGAAAGATTGACATGCAATTTTCATCTTCTGGGGAAAAAAGCGCTTATCGTCAG 129
QY 34 ArgAlaAlaAsnGlyGluTrpArgAlaIleSerTyrAlaGluMetPheHisValAlaArg 53
DB 130 TGGAGGCTTAAGTCTAA---ATCGCCATGCTATGGCGCTGACTACCAAGAAACACGT 186
QY 54 AlaIleAlaGlnSerLeuLeuProTyrGlyLeuSerAlaGluArgProLeuLeuVal 73
DB 187 GCATCTCAATTCATTTACTCATATGGCGTGGCGCTTCAAGAAAAATGGGGGATTTT 246
QY 74 SerGlyAsn-----AspLeuGlnHisLeuGlnLeuAlaPheGly 86
DB 247 TCCCAAAAGACATGATTGCTATGCTATGCTATGATTAATGCTACTTGCAGTTGGCT 300
QY 87 AlaMetTyrAlaGlyLysProTyrCysProValSerProAlaTyrSerLeuLeuSerGln 106
DB 301 -----GCTGCACAGTTCCTTTATTCACCAAGTACTGTCG 336
QY 107 AspLeuAlaLysLeuArgHisIleValGlyLeuLeuGlnProGlyLeuValPheAlaAla 126
DB 337 GAACAAGCC-----GCTTATATCATCAACGATCGGATATTCGTATCTTTTGTGGT 390
QY 127 AspAlaAlaProPheGlnArgAlaIleGlu----- 136
DB 391 GACCAAAAGAAATATGACGTTGTTCTGAGTTAAGCGCTATTATGCCGCACTTAACAT 450
QY 137 -----ThrIleLeuProAspAspValProAlaIlePheTrpArg 149
DB 451 ATTATGTTGTTAATTTCTCAGGCTGTTTAAATCCAAACATACCTTTATCATCTGACA 510
QY 150 GlyLysLeuAlaGlyArgArgThrValSerPheAspSerLeuLeuGlnProGlyGly 169
DB 511 AACTTATACAGTACGCTCAATCT---CAATAGATCTCTGCTACACAA----- 558
QY 170 IleGluAlaAspAsnAlaPheAlaAlaThrGlyProAspThrIleAlaLysPheLeuPhe 189
DB 559 ---CGTATGAGAAATGTTT-----GATGATTTGTTTACAGATTTAT 603
QY 190 ThrSerGlySerThrLysLeuProLysAlaValProThrThrGlnArgMetLeuCysAla 209
DB 604 ACCTCGGTACAAACAGGTGAGCTTAAGGCTGTTATGCTGATTAACCAATCTGGCTCT 663
QY 210 AsnGlnGlnMetLeuLeuGlnThrPheProValPheGlyGluGluProProValLeuVal 229
DB 664 -----CAACTTATGCTATGATGATCAAGCGCTTTCATTAATCGGATTAAGATGTTTCATTA 717
QY 230 AspTrpLeuProTrpAsnHisThrPheGlyGlySerHisAsnIleGlyIleValLeuTyr 249
DB 718 TGGTTTATCCGTTATCAATGATTTGAACGGCATGAGCTTCAATGATTAATGCTACT 777
QY 250 AsnGlyGlyThrTyrTyrLeuAspArgGlyLysProThrAlaGlnGlyPheAlaGluThr 269
DB 778 GGCGCTATTAAATGTTATCTCACAGTACCCTAGCCGTAAGGAAGCGATGCT----- 831
QY 270 LeuArgAsnLeuSerGlnLysSerProThrAlaTyrLeuThrValProLysGlyTrpGlu 289
DB 832 -----GCAGTAAACCAACGCTTATGCTGACAGTGCAGCTTTTACGAA 876
QY 290 GluLeuValAlaValLeuGlnArgAspSerThrLeuArgGluArgPhePheAlaArgMet 309
DB 877 AAAGTATATAGCGCATTCACAGAAAGTTTCAAGACCTCGACTA-----CGCCAA 930
QY 310 LysLeuPhePhePheAla-----AlaAlaGly 318
DB 931 TCGATGTTTAAATGGCTCTTAAACAGGCTGAAAAACAAACGTCAAGCACATTAAGCTGGA 990

```


LOCATION: (47036) .. (47036)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (51334) .. (51334)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (51662) .. (51662)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (51786) .. (51786)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (51805) .. (51805)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (55369) .. (55369)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (65309) .. (65309)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (65313) .. (65313)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (80024) .. (80024)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (100091) .. (100091)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (102696) .. (102696)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (105121) .. (105121)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (107248) .. (107248)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (117136) .. (117136)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (119750) .. (119750)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (119924) .. (119924)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (120038) .. (120038)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (121344) .. (121344)	
OTHER INFORMATION: n equals a, t, g or c	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (122167) .. (122167)	

Pred. No.: 1,76e-16 Length: 1830121
 Score: 283.00 Matches: 137
 Percent Similarity: 38.44% Conservative: 114
 Best Local Similarity: 20.98% Mismatches: 252
 Query Match: 9.18% Indels: 150
 DB: 4 Gaps: 28

US-09-750-986d-30 (1-589) x US-09-643-990A-1 (1-1830121)

QY 15 LeuGluArgLeuGluHisThrPalaIleThrArgProGluGlnThrCysValAlaIleArg 34
 Db 646 GTTCATGATTTTCAACAACAGCCAAATCTGTCAAAATATGACCGCATTCGTTATATA 705
 QY 35 AlaAlaAsnGluThrPalaArgIleSerTyraIleGluMetPheHisAsnValArgAla 54
 Db 706 ---GAACACGCGCTTGCGGAGATATCTTTGAAAAAATTTCAGAGCAACTCAATCAA 762
 QY 55 IleAlaGlnSerLeuLeuProTyrglyLeuSerAlaGluArgProLeuIleValSer 74
 Db 763 CTTTCTCGACATTTGCTGCTCAAAATATGACGTACAAATAAATCGCATTTTGGCC 822
 QY 75 GlyAsn-----AspLeuGlnHisLeuGlnLeuAlaPheGlyAla 87
 Db 823 CATTAATTTGGAAGCTTGACAAATCTTGACATTCGACCTTACAAATTCGA----- 873
 QY 88 MetTyraIleGlyIleProTyrglyProValSerProAlaTyreSerLeuLeuSerGlnAsp 107
 Db 874 -----GCAATCACAGTGCCTATTTCACGACCAACCAATACAGCCGAG 912
 QY 108 LeuAlaIleLeuArgHisIleValGlyLeuLeuGlnProGlyLeuValPheAlaAsp 127
 Db 913 CAAGCGAATTT-----ATCTTAATACAGCGCATTAATAAATCTTCTGTCGCGGAT 966
 QY 128 AlaAlaProPheGlnArgAlaIleGluThrIleLeuProAspAspValProAlaIlePhe 147
 Db 967 CAAGAGCAATACGATCAACATTCGAA-----ATTCTCATCTTGTCCAAATTA--- 1017
 QY 148 ThrArgGlyIleLeuAlaGlyArgArgThrValSerPheAspSerLeuLeuGlnPro 167
 Db 1018 -----CAAAAAATTGTACATGAATCCACCATTTCAA----- 1050
 QY 168 GlyGlyIleGluAlaAsp-----AsnAlaPheAlaIleThrGlyPro 181
 Db 1051 -----TTACAAACAGATCTCTTTCTTGACCTTGAGAAAGTTTATTAAACAGGTTCA 1104
 QY 182 AspThr-----IleAla 185
 Db 1105 AACGCCCAACAGATGAATTAACCAACGCTTAAACCAAAAAACATTAATGCGATTATTT 1164
 QY 186 LysPheLeuPheThrSerGlySerThrLysLeuProLysAlaValProThrThrGlnArg 205
 Db 1165 ACATATTATTATACCTCAGGCACACAGCGAGAGCTTAAGGTGCATGTTAGATTACGCT 1224
 QY 206 MetLeuCysAlaAsnGlnMetLeuLeuGlnThrPheProValPheGlyGluGlnPro 225
 Db 1225 AATCTGCT-----CACCAATTAGAACAACATGATCTTCACTTAATGTGACAGATCAG 1278
 QY 226 ProValLeuValAspThrLeuProTyraHisThrPheGlyGlySerHisAsnIleGly 245
 Db 1279 GATATTTCATCTCTTTTAACTCTCTCATATTATTAAACGGGATGG---GCCGCT 1335
 QY 246 IleAlaLeuTyraAsnGlyThrTyra---TyreLeuAspAspGlyLysProThrAlaGln 264
 Db 1336 TATATTCTTCATAGAGCGGCAATACCTTGTGATTAGAAAGACATATCAAGTCGGCTCA 1395
 QY 265 GlyPheAlaGluThrLeuArgAsnLeuSerGlnIleSerProThrAlaTyreLeuThrVal 284
 Db 1396 GCT-----TTAACGAAATTCGCCCACTTAATGTGCGCGTA 1434
 QY 285 ProLysGlyTyrrpGluLeuValGlyAla-----LeuGluArgAspSerThr 300
 Db 1435 CCACGTTTTCAGAAAAAATTATGCTGCGGTATGGAATAAAGTTCAAAAAAGCCCAAAA 1494

QY 301 LeuArgGluThrPhePheAlaArgMetLysLeuPhePheAlaIleAlaGlyLeuSer 320
 Db 1495 CTTGCGCA-----ATTATGTTTCATTGGCAATTCGTTGGGA 1533
 QY 321 GlnGlyIleThrPhe---ArgLeuAspArgValAla----- 331
 Db 1534 CAAAAATATTTTATCTACGTGCGAATTAACAAAGCTATTCGTTTATTGAAGAAACAA 1593
 QY 332 -----GluGlnHisCysGlyGluArgIle 339
 Db 1594 TTTCCTTAGACATTAATTAATGCTCTCAAAACTGCTCATTAATTATGGGGAGCAATA 1653
 QY 340 ArgMetMet----- 342
 Db 1654 AAATGATGCTTGCGGAGGACCAAAATTGAACCGGCTATTTGGGCTATTTTCATGCT 1713
 QY 343 -----AlaGlyLeuGlyMetThrGlnThrAlaPro-----SerCys 354
 Db 1714 ATTGATATCAACATCAAAATTAAGCTATGTATGACAGAAACAACTGCAACCGTTCTTC 1773
 QY 355 -----ThrPheThrThrGlyProLeuSerMetAlaGlyTyrrIleGlyLeuProAlaPro 372
 Db 1774 TGGCATGATTTCCAAATTAACCAAAATTC-----ATCGGACACATGATGCCA 1821
 QY 373 GlyCysGluValIleLeuValProValAspGlyLysLeuGluGlyArgPheHisGlyPro 392
 Db 1822 AAAGCGAAGTGAAA-----ATTGGGAAATATGAAATCTTGTGCGTGGCGGA 1872
 QY 393 HisValMetSerGlyTyrrTPAArgAlaProGluGlnAsnAlaGlnAlaPheAspGluGlu 412
 Db 1873 ATGGTATGAAAGCTATTAACAGAGCCAGAAAGAAACGGCTCAAGCTTACCGAAGAT 1932
 QY 413 GlyTyrrCysSerGlyAspAlaIleLysLeuAlaAspProAlaAspProGlnLysGly 432
 Db 1933 GGTTCCTTAATAAAGCTGCGATGCAAGAGATTT-----GACGAACAAGCAAT 1980
 QY 433 LeuMetPheAspGlyArgIleAlaGluAspPheLysLeuSerSerGlyValPheValSer 452
 Db 1981 TTAATTATTCATCGATGCTATCAAAAGATTAATGAAGAACTCAACGGCAATATATGCA 2040
 QY 453 ValGlyProLeuArgThrArgAlaValLeuGluGlyGlySerTyrrValLeuAspValVal 472
 Db 2041 -----CCACAAATATATCGAAGCAAAATCGTAAAGATTAATTAACAAATTCGCG 2094
 QY 473 ValAlaAlaProAspArgGluCysLeuGlyLeuLeuValPheProArgLeuLeuAspCys 492
 Db 2095 ATCATGCTGATCGGAAAAATATGTATCCGCGCTTATTTGCT-----TGC 2142
 QY 493 ---ArgAlaLeuSerGlyLeuGlyLysGluAlaSer-----AspAlaGlu 506
 Db 2143 TTGTATAGTTTGAAGAAATACGCTTAACAGCTCAATTAATATCATGACCGTTAGAA 2202
 QY 507 ValLeuAlaSerGluProValArgAlaIlePheAlaAspThrLeuLysAlaGluAsnArg 526
 Db 2203 CTACTAAAAATTCGACATCTGAAATGTGTGAGCATGTATTATAGCGGTGCAAAA 2262
 QY 527 GluAlaThrGlyAsnAlaSerArgIleMetTyrrValGlyLeuLeuAspThrProPheSer 546
 Db 2263 GAA---TTGGCTCATTTTCGACAGAGTAAAAAATTCACGTTACTTTCTCAAGATTACG 2319
 QY 547 IleAspLysGlyLysValThrAspLysGlySerIleAsnGlnArgAlaValLeuGlnTrp 566
 Db 2320 ATTAATTTAGGCAATTTACCAACATTAATAATTCGTGAAGAAAGATTTTAGAAGCT 2379
 QY 567 ArgSerAlaLysValAspAlaLeuTyrrArgGlyGluAsp 579
 Db 2380 TATCGCAAGCAAAATTAAGCAATGTATCATTCACAAAGAA 2418

RESULT 8
 US-10-329-960-1
 ; Sequence 1, Application US/10329960
 ; Patent No. 6742927
 ; GENERAL INFORMATION:

```

Db 1105 AACGCCAACAGATGACTAACCAACGCTTAACCAAAAAACATTAATGATTTATT 1164
Qy 186 LysPheLeuPheTherSerGlySerThrLysLeuProLysAlaValProThrThrGlnArg 205
Db 1165 AGCATTAATTAATCTCCAGGACACAGCGGAGAGAGCTTAAGGTCTCATGTTAGATTACGCT 1224
Qy 206 MetLeuCyAlaAsnGlnMetLeuGlnThrPheProValPheGlyGluPro 225
Db 1225 AATCTGCT-----CACCAATTAGAAACACATGATCTTCACTTAATGACAGATCAG 1278
Qy 226 ProValLeuValAspTrpLeuProTrpAsnHisThrPheGlySerHisAsnIleGly 245
Db 1279 GATATTTCATCTTTTACCATCTCTCATATTATTGAAGGGGATGG---GGGGCT 1335
Qy 246 IleValLeuTrpAsnGlyGlyThrTyr---TyrLeuAspAspGlyLysProThrAlaGln 264
Db 1336 TAAATCTTCATGAGCGGCAATACCTTGTCTATTGAGACACTATCAAGTCGGGTCA 1395
Qy 265 GlyPheAlaGluThrLeuArgAsnLeuSerGluIleSerProThrAlaTyrLeuThrVal 284
Db 1396 GCT-----TTAACGAAATTCGCCCACTTAATGTCGCCGCTA 1434
Qy 285 ProLysGlyTrpGluGluLeuValGlyAla-----LeuGluArgAspSerThr 300
Db 1435 CCACGTTTATGAGAAAAAATTATGCTGCCGCTATTGGATTAAGTCAAAAAAGCCCAAA 1494
Qy 301 LeuArgGluArgPhePheAlaArgMetLysLeuPhePheAlaAlaAlaGlyLeuSer 320
Db 1495 CTTCGCCAA-----ATTATGTTTCATTGGGCAATTCCTGGGA 1533
Qy 321 GlnGlyIleTrpAsp---ArgLeuAspArgValAla----- 331
Db 1534 CAAAAATATTTTATCTACGTCGCAATTAACAAAGCTATTCCTCTTATTGAAGAAACAA 1593
Qy 332 -----GlnGlnHisCysGlyGluArgIle 339
Db 1594 TTGCTTTAGACATTAATTAATGCTCTCAAAACCTTCGCAATTATGGGGGACGAAATA 1653
Qy 340 ArgMetMet----- 342
Db 1654 AAAATGATGCTTCGAGAGAGCAAAATTGAACCGGCTATTGGGCTATTTTCCATGCT 1713
Qy 343 -----AlaGlyLeuGlyMetThrGluThrAlaPro-----SerCys 354
Db 1714 ATTGGATTCACAACTCAAAATTAGGCTATGATGACAAACCACTGACACCTTTCTTGC 1773
Qy 355 -----ThrPheThrThrGlyProLeuSerMetAlaGlyTyrIleGlyLeuProAlaPro 372
Db 1774 TGGCATGATTTCCAAATTTAACCAAAATTCA-----ATCGCACACATGATGCCA 1821
Qy 373 GlyCysGluValLysLeuValProValAspGlyLysLeuGluGlyArgPheHisGlyPro 392
Db 1822 AAAGCGGAAGTGA---ATTCGGGAAATTAATGAATCCTTGTGCGTGGCGGA 1872
Qy 393 HisValMetSerGlyTyrTrpArgAlaProGluGlnAsnAlaGlnAlaPheAspGluGlu 412
Db 1873 ATGGTATGAAGGCTTTTCAAGAACCCGAAAGAAACGCTCAAGCTTACACGAAAGAT 1932
Qy 413 GlyTyrTyrCysSerGlyAspAlaIleLysLeuAlaAspProAlaAspProGlnLysGly 432
Db 1933 GGTTCCTTAATAAATCGCGCATGACAGGAATTT-----GACGAACAAGGCAT 1980
Qy 433 LeuMetPheAspGlyArgIleAlaGluAspPheLysLeuSerSerGlyValPheValSer 452
Db 1981 TTTTATTATTAACGATGCTATCAAGAAATTAATGAACCTCAACGCAATATATTCGCA 2040
Qy 453 ValGlyProLeuArgThrArgAlaValLeuGluGlyGlySerTyrValLeuAspValVal 472
Db 2041 -----CCACAAATATATCGAAACCAAAATCGTAAAGATAAATTATCGAAACAATTGGC 2094
Qy 473 ValAlaAlaProAspArgGluCysLeuGlyLeuLeuValPheProArgLeuLeuAspCys 492
Db 2095 ATCATCGCTGATCGRAAAATATATGATCCGCGCTTATGTGCT-----TGC 2142

```

```

Qy 493 ----ArgAlaLeuSerGlyLeuGlyLysGluAlaSer-----AspAlaGlu 506
Db 2143 TTGATAGTTAGAGAAATACCGTTAAACAGCTCAATATTAAATATATCATGACGCTTAGAA 2202
Qy 507 ValLeuAlaSerGluProValArgAlaTrpPheAlaAspTrpLeuLysArgLeuAsnArg 526
Db 2203 CTACTTAAAAAATTCGACATTCCTGAAAAATGTTTGACACATCGTATTAAATGGGTGCAAAA 2262
Qy 527 GluAlaThrGlyAsnAlaSerArgIleMetTrpValGlyLeuLeuAspThrProProSer 546
Db 2263 GAA---TTGGCTCATTTCCAGCAAGTAAAAAAAATTCACCTTACTTCTCAAGCATTCAGC 2319
Qy 547 IleAspLysGlyGluValThrAspLysGlySerIleAsnGluArgAlaValLeuGlnTrp 566
Db 2320 ATTAATTAATGAGCGCAATTCACCAACACATTAATTAATTCGTAGAAAAGTATTTAGAACGT 2379
Qy 567 ArgSerAlaLysValAspAlaLeuTyrArgGlyGluAsp 579
Db 2380 TATCGCAAGCAATTAAGCAATGATCATTCACAAAGAA 2418

RESULT 7
US-09-643-990A-1
; Sequence 1, Application US/09643990A
; Patent No. 6528289
; GENERAL INFORMATION:
; APPLICANT: Robert D. Fleischmann
; Mark D. Adams
; Owen White
; Hamilton O. Smith
; J. Craig Venter
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; thereof, and Uses Thereof
;
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville,
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; OPERATING SYSTEM: MS DOS v6.22
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/643,990A
; FILING DATE: 23-Aug-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,429
; FILING DATE: 1995-06-07
; APPLICATION NUMBER: 08/426,787
; FILING DATE: 1995-04-21
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB186P1C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-610-5790
; TELEFAX: 310-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830121 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
; US-09-643-990A-1
Alignment Scores:

```

QY		372	ProG1XySpslValVlylsleu-----ValProValAaprgLylys	384
Db		33840	GCGGGCCTGCAGACACCAAGCTGTTCGATGAATCCGGAGTCCACTGCCCTCACAGCGGGAA	33899
QY		385	LeuglUglYarPhe-----HiglyProHisValwetsergLYTyrrTrparGalapro	402
Db		33900	ACCGTTGGGAGGCTTCAGTTCCTAGTCGACGATGTTCGGCGCATTAATGACCG	33958
QY		403	GlUglnAsnAlaglnAlpheApnglUglTYTyrrCysserGLylasphAlaleys	422
Db		33960	GAAGCCGACACCAGCGGCTTGACAGAGACGGTTGTGACCTGACAGTAT--GTTGCG	34016
QY		423	LeuAlaAspProAlaAspProGlnLysGLyleuMetPheApriYargilealaglueAr	442
Db		34017	GTCGTGACAGATGGCGCGGATGATCCG-----ATCGTGGTCCGAGTCCGTTGAC	34066
QY		443	PheIylseuSerserGLylvalPhevalSerValGLylProleuarGrThraAlavalLeu	462
Db		34068	CtAATCAATTGGCGCGATACCGGATCGCGCAGAGAAATC-----GAGATGGCGCTG	34120
QY		463	GlUglYglYserTyrrValLleuaspyAl---ValValAlaAlproAspargrGucysleu	481
Db		34122	CtAGAGATATCCGAGATGTGCGGAGGTGGCAGTTGTGCGGCTGCTTAC--GAGGATTTG	34178
QY		482	GLYleuLeuValPheProArgleuLeuAspCYaArgAlaleuserglyleuglyLysGlu	501
Db		34179	GGTCAACGATCGTGTGATTCGTC-----GTAGCGCAGAG	34214
QY		502	AlaserAspAlaclValLleualaser-----Glu	511
Db		34215	GCAATTAAATCCCGATGATGATTAACTATGTTGCCAACAGCTTTCGATTAACGCG	34274
QY		512	ProValarglArThrPheAlaAspTPbleuYargleuAsnArgGlnAlathrGlyAsn	531
Db		34275	CCGGTGAAGTCCGCTTCGTGACGCGGCTGCCGCG--AAGC--GATGGAAAAGTCT	34330
QY		532	AlaserArgileuMetTPvalGLyleuLeuAspThrProPserlleAsplysGlyGlu	551
Db		34331	CAAAAAGCACTTGTCTCTTCGAGGAGCTGACGCGGCTACACGACGATGATTAATGCGGCG	34390
QY		552	ValThrAspIysglYserlleasnGlnArgAlaValleuglnTP-----	566
Db		34391	ACGCACTCTTTGGACCGGACGCGGCGGATGTGTGGTTGGTCGCGCGCCACTGT	34450
QY		567	ArgSerValIasValAspAlaleuTyrr-----ArgGLyluasApGlnser	581
Db		34451	AGAGCTTCAAAGGACACGCGCAATGTGACCTGATGTGTGCGAGGGGTGCGCGGTAGAAGT	34510
QY		582	MetLeuArgAsp	585
Db		34511	TTTTACGCGAC	34522
RESULT 6				
US-09-557-884-1				
Sequence 1, Application US/09557884				
Patent No. 6506581				
GENERAL INFORMATION:				
APPLICANT: Fleischmann et al.				
TITLE OF INVENTION: The Nucleotide sequence of				
the Haemophilus influenzae Rd Genome, Fragments				
Thereof, and Uses Thereof				
NUMBER OF SEQUENCES: 1				
CORRESPONDENCE ADDRESS:				
ADDRESSEE: Human Genome Sciences, Inc.				
STREET: 9410 Key West Avenue				
CITY: Rockville				
STATE: MD				
COUNTRY: USA				
ZIP: 20850				
COMPUTER READABLE FORM:				
MEDIUM TYPE: 3 1/2 inch diskette				
COMPUTER: Dell Pentium				

```

1      OPERATING SYSTEM: MS DOS V6.22
2      SOFTWARE: ASCII Text
3      CURRENT APPLICATION DATA:
4      APPLICATION NUMBER: US/09/557, 884
5      FILING DATE: 25-Apr-2000
6      CLASSIFICATION: <Unknown>
7      PRIOR APPLICATION DATA:
8      APPLICATION NUMBER: 08/476, 102
9      FILING DATE: JUN-5-1995
10     ATTORNEY/AGENT INFORMATION:
11     NAME: Michelle S. Marks
12     REGISTRATION NUMBER: 41, 971
13     REFERENCE/DOCKET NUMBER: PB186P3
14     TELECOMMUNICATION INFORMATION:
15     TELEPHONE: 301-309-8504
16     TELEFAX: 301-309-8439
17     INFORMATION FOR SEQ ID NO: 1:
18     SEQUENCE CHARACTERISTICS:
19     LENGTH: 1830121 base pairs
20     TYPE: nucleic acid
21     STRANDEDNESS: double
22     TOPOLOGY: linear
23     SEQUENCE DESCRIPTION: SEQ ID NO: 1:
24     US-09-557-884-1

```

Alignment Scores:			
Pred. No.:	1,76e-16	Length:	1830121
Score:	283.00	Matches:	137
Percent Similarity:	38.44%	Conservative:	114
Best Local Similarity:	20.98%	Mismatches:	252
Query Match:	9.18%	Indels:	150
DB:	4	Gaps:	28
US-09-750-986D-30 (1-589) x US-09-557-884-1 (1-1830121)			
Qy	15	LeuGLuArGLeuGLuHISrTPAlALuYThrArGProGLuGLuThrCyVaLaLaLaArG	34
Db	646	GTTCATGCGTATTCACAAACAAGCCAAACTGTCMAATATGACGCGCATTCGCTATGAA	705
Qy	35	AlaLaLaAnGlyGLuTPrArGArgLLeSerThyAlaGLuMetPheLaAsnValArgLa	54
Db	706	---GAACACGGCTGTGGCCGAGATATCTTGGAAAACTTCAAGACCAATCATGCA	762
Qy	55	IleAlaGLnSerLeuLeuProThyGlyLeuSerAlaGLuArGProLeuLeuIleValSer	74
Db	763	CTTTCGTGAGCATTCGTTGCTCACAAATATGACGTCACAAAGATAAATGCGCATTTTCC	822
Qy	75	GlyAsn-----AspLeuGLuHISLeuGLnLeuAlaPheGLyAla	87
Db	823	CATATATATGCAACGTGGACAAATCGTTGACATTCGACCTTACAAATTGCA-----	873
Qy	88	MetThyAlaGLyLLeProThyCyseProValSerProAlaThySerLeuLeuSerGLnAsp	107
Db	874	-----GCATCAACAGTGCCTATTTCGCAACCAATACAGCCAG	912
Qy	108	LeuAlaLysLeuArgHISileValGlyLeuLeuGLnProGlyLeuValPheAlaLaAsp	127
Db	913	CAACGACAGATT-----ATCCTAAATCAGCCGATGTAAAAATCTCTTCGTGGCGCAT	966
Qy	128	AlaLaLaProPheGLnArgLaIleGLuThrLLeuProAspValProAlaIlePhe	147
Db	967	CMAAGCAATATCGATCAAAACATTGGAA-----ATTGCTCATCTTGTCCAAATTA---	1017
Qy	148	ThyArgGLyGLuLeuAlaGLyArGArgThyValSerPheAspSerLeuLeuGLnPro	167
Db	1018	-----CAAAAAATTATGCAATGAATACACCATTCMA-----	1056
Qy	168	GlyGLyLLeGLuAlaAsp-----AsnAlaPheAlaLaThyGlyPro	181
Db	1051	-----TTACACAAAGATCCCTTTCTTGACCTTGGGAAAGTTTATTAACAGGTTCA	1104
Qy	182	AspThr-----IleAla	185

Db 18143 GCATTAGATCCGATGATTGTAATGTTGCCCAACAGCTTCGATACATAAGCCG 18202
Qy 512 Pro-ValArgAlaTrpPheAlaAspTrpLeuValArgLeuAsnArgGluAlaThrGlyAs 531
Db 18203 CCGCGCGAAGTCC-----GCTTCGTGAGACGCCCTGCCCAACGCGATGGGAAGTCC 18256
Qy 531 nAlaSerArgIleMetTrpValGlyLeuLeuAspThrProProSerIleAspGlyGly 551
Db 18257 TCAAAAAGCAGTTCCTTCGAGAGGCTGACTGCCGCTACACAGATGATGATATGCGGG 18316
Qy 551 uValThrAspLysGlySerIleAsnArgAlaValLeuGlnTrp-----566
Db 18317 CACGCATGCTTGCACCGGACGGGGCCGATGTTGTTGGTCCGCCGCACTG 18376
Qy 567 -ArgSerAlaLysValAlaAspAlaLeuTy-----ArgGlyGluAspGlnSe 581
Db 18377 TGAAGCTTCAGAGCCACCGCATGTGACAGTGTGTGTGAGGGGTCCGCTAGAG 18436
Qy 581 rMetLeuArgAsp 585
Db 18437 TTTTTCAGCGGAC 18449

RESULT 5
US-08-311-731A-138
Sequence 138, Application US/08311731A
Patent No. 6583266
GENERAL INFORMATION:
APPLICANT: SMITH, DOUGLAS
APPLICANT: MAO, JEN-I
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR
NUMBER OF SEQUENCES: 411
CORRESPONDENCE ADDRESS:
ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
STREET: 600 ATLANTIC AVENUE
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Releasee #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/311, 731A
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: GATES, EDWARD R.
REGISTRATION NUMBER: 31,616
REFERENCE/DOCKET NUMBER: C0044/7125
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/720-3500
TELEFAX: 617/720-2441
INFORMATION FOR SEQ ID NO: 138:
SEQUENCE CHARACTERISTICS:
LENGTH: 35961 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Mycobacterium leprae
US-08-311-731A-138
Alignment Scores:
Pred. No.: 1,78e-21 Length: 35961
Score: 302.50 Matches: 150
Percent Similarity: 37.50% Conservative: 69

Best Local Similarity: 25.68% Mismatches: 234
Query Match: 9.82% Indels: 132
DB: 4 Gaps: 27
US-09-750-986D-30 (1-589) x US-08-311-731A-138 (1-35961)
Qy 44 SerTyAlaGluMetPheHisAsnValArgAlaIleAlaGlnSerLeuProTyGly 63
Db 33039 TCGGTGCTGAGCGGTTCGAGAGTGCACCGCTGTGAGCGGTGTGCGCCGCGCCG 33092
Qy 64 LeuSerAlaGluArgProLeuLeuIleValSerGlyAsnAspLeuGlnIleuGlnLeu 83
Db 33093 ---ACGCATTCGACCTGTGATGAGATTACT-----33119
Qy 84 AlaPheGlyAlaMetTyAlaGlyIleProTyTyCysProValSerProAlaTySerLeu 103
Db 33120 ---GGTTCCTGATGTCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 33173
Qy 104 LeuSerGluAspLeuAlaLysLeuArgHisIleVal-----GlyLeu 117
Db 33174 GTCGAACG-----CGCATATCTACACCGATTCGCGCGCGCGCTCG 33218
Qy 118 LeuGlnProGlyLeuValPheAlaAlaAspAlaIleProPheGlnArgAlaIleGluThr 137
Db 33219 CTGCGTCCGCGGCGCC---AGTCCGACCTCGCGC---33248
Qy 138 IleLeuProAspAspValProAlaIlePheThrArgGlyGluLeuAlaGlyArgArgThr 157
Db 33249 ---CCGACCGGTGCGCGCATATC-----CCG 33272
Qy 158 ValSerPheAspSerLeuLeuGlnProGlyGlyIleGluAlaAspAsnAlaPheAla 177
Db 33273 GTTCACTAGATCT-----CGCTCCGAAACCGCTATCT 33308
Qy 178 AlaThrGlyProAspThrIleAlaLysPheLeuPheThrSerGlySerThrLysLeuPro 197
Db 33309 GACCGCTCACCTGATCCACCGGATGATGATGATGATGATGATGATGATGATGATG 33368
Qy 198 LysAlaValProThrThrGlnArgMetLeuCyAlaAsnGlnIleMetLeuGlnThr 217
Db 33369 AAGGGGTACTGTGACCGCGGACGAGGATCGCGGTGATTTGACCGCTTCCGCGGCC 33428
Qy 218 PheProValPheGlyGluGluProProValLeuValAspTrpLeuProThrPheHisThr 237
Db 33429 TGGCAGTGCAGCGCCGATGAC-----GTTTGATGATGATGATGATGATGATGATG 33482
Qy 238 PheGly-----GlySerHisAsnIleGlyIleValLeuTyraengly 251
Db 33483 CACGGCTTGATCGGTGCTTGTGATGCTCGGATT-----33521
Qy 252 GlyThrTyTrpLeuAspAspGlyLysProThrAlaGlnGlyPheAlaGluThrLeuArg 271
Db 33522 GGAATCGCTTGTACACCGGAAACCCACACCGACCGCTATGCTCAGGCT-----33575
Qy 272 AsnLeuSerGluIleSerProThrAlaTyLeuThrValProLysGlyTrpGluGluLeu 291
Db 33576 ---TGCTGTGAGCGCGGGGATCTCTGTATTTGCGGCTTCCACGATGCTGCGGCTG 33632
Qy 292 ValGlyAlaLeuGluArgAspSerThrLeuArgGluArgPheAlaArgMetLysLeu 311
Db 33633 GTGCA-----GACGAGCGAGTT---GCCCGGCGCTTAAGACCGCGCGGCTA 33677
Qy 312 PhePhePheAlaAlaAlaGlyLeuSerGlnGlyIleTrpAspArgLeuAspArgValAla 331
Db 33678 CTAGTCTGAGGAGTGCCTCATTTGCGGTTCGCTTGTGATGATGATGATGATGATGATG 33737
Qy 332 GluGlnHisCysGlyGluArgIleArgMetMetAlaGlyLeuGlyMetThrGluThrAla 351
Db 33738 GGTCACTCGGCCCATCTCAACG-----TATGCGACGACGGAATCAGTG 33779
Qy 352 ProSerCysThrPheThrThrGlyProLeuSerMetAlaGlyTyrlleGlyLeuProAla 371
Db 33780 ATCACCTGTGACGTTGCGGACGCGTGAAGCGCGCGGTGCGGTGCGGTGCGGTGCGGTG 33839

TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
 NUMBER OF SEQUENCES: 411
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: WOLF, GREENFIELD & SACKS, P. C.
 STREET: 600 ATLANTIC AVENUE
 CITY: BOSTON
 STATE: MASSACHUSETTS
 COUNTRY: USA
 ZIP: 02210
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/311,731A
 FILING DATE:
 CLASSIFICATION: 530
 ATTORNEY/AGENT INFORMATION:
 NAME: GATES, EDWARD R.
 REGISTRATION NUMBER: 31,616
 REFERENCE/DOCKET NUMBER: C0044/7125
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 617/720-3500
 TELEFAX: 617/720-2441
 INFORMATION FOR SEQ ID NO: 24:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 38494 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: circular
 MOLECULE TYPE: DNA (genomic)
 HYPOTHEtical: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: MYCOBACTERIUM LEPRAE
 US-08-311-731A-24

Alignment Scores:
 Pred. No.: 3,71e-22 Length: 38494
 Score: 309.00 Matches: 149
 Percent Similarity: 37.78% Conservative: 72
 Best Local Similarity: 25.47% Mismatches: 231
 Query Match: 10.03% Indels: 133
 Gaps: 27

US-09-750-986d-30 (1-589) x US-08-311-731A-24 (1-38494)

QY 44 SerTyrAlaGluMetPheHisAsnValArgAlaIleAlaGlnSerLeuProTyrGly 63
 DB 16967 TCGGTGCTGAGCGGTGTGCGAGGTGCACGCTGTGTAGCGGTGCTGCCACGCCG 17020
 QY 64 LeuSerAlaGluArgProLeuLeuIleValSerGlyAsnAspLeuGlnHisLeuGlnLeu 83
 DB 17021 ---ACGGCATCGACTGTGTCTAGCGATTACT- 17047
 QY 84 AlaPheGlyAlaMetTyrAlaGlyIleProTyrCysProValSerProAlaTyrSerLeu 103
 DB 17048 ---GTTGGCTGATTCGCGGTGTGCGCGTGTGCTGCCGCGCATTCGCGTGTG 17101
 QY 104 LeuSerGlnAspLeuAlaIleValSerGlyHisIleVal- 117
 DB 17102 GTTCGACGCG- 17146
 QY 118 LeuGlnProGlyLeuValPheAlaIleAspAlaIleProPheGlnArgAlaIleGluThr 137
 DB 17147 CTGGGTCCGGGGCC- 17176
 QY 138 IleLeuProAspAspValProAlaIlePheThrArgGlyGluLeuAlaGlyArgThr 157
 DB 17177 ---CCGGACCGGTTCGCGCATTC- 17200
 QY 158 ValSerPheAspSerLeuLeuGlnProGlyGlyIleGluAlaAspAlaPheAla 177

DB 17201 GTTCACCTAATGCT- 17236
 QY 178 AlaThrGlyProAspThrIleAlaIlePheLeuPheThrSerGlySerThrIleLeuPro 197
 DB 17237 GAGCCGCTGATGATGACCGCGATGTGATCTACACCTCGGATCTACGGGCCCCCT 17296
 QY 198 IysAlaValProThrGlnArgMetLeuCysAlaAsnGlnMetLeuLeuGlnThr 217
 DB 17297 AAGGGGTACTGTGACCGCGCGATGCGCGTGTATGACCGCGCTTGGCCAGGCC 17356
 QY 218 PheProValPheGlyGluGlnProProValLeuValAspThrLeuProThrPheHisThr 237
 DB 17357 TGGCAGTGAACGCGCGATGAC- 17410
 QY 238 PheGly- 251
 DB 17411 CAGGCTTGTACTCGGTTTGTGATCTCGGAT- 17449
 QY 252 GlyThrTyrTyrLeuAspAspGlyLysProThrAlaGlnIlePheAlaGluThrLeuArg 271
 DB 17450 GGAATCGCTTGTGACACCGGAAACCCACTCCGACCGCTATGCTCAGGCT- 17503
 QY 272 AsnLeuSerGluLeuSerProThrAlaTyrLeuThrValProLysGlyTyrGluGlnLeu 291
 DB 17504 ---TGCTGTGAACCGGGGATCTGTATTCGCGCTTCCACCGATGTGCGGCTG 17560
 QY 292 ValGlyAlaLeuGlnArgAspSerThrLeuArgGluArgPhePheAlaArgMetLysLeu 311
 DB 17561 GTGGCA- 17605
 QY 312 PhePhePheAlaAlaAlaGlyLeuSerGlnGlyIleThrAspArgLeuAspArgValAla 331
 DB 17606 CTAGTGTCTGGAGTGCTGATTCGCGGTCCGATTTGATGCGGTAGCGACCTCACA 17665
 QY 332 GlnGlnHisCysGlyGluArgIleArgMetMetAlaGlyLeuGlnMetThrGluThrAla 351
 DB 17666 GGTTCATCGGCCCATCGAACCG- 17707
 QY 352 ProSerCysThrPheThrThrGlyProLeuSerMetAlaGlyTyrIleGlyLeuProAla 371
 DB 17708 ATCACTGTGTGACGTTGGCGGACCGGTGACCGCGCTGCGGTGCGGATTCGCGCTG 17767
 QY 372 ProGlyCysGlyValLysLeu- 384
 DB 17768 GCGGGCGTGACACACGAGCTGTGATGATCCGAGTCCAGTGCCTTACGAGGGGAA 17827
 QY 385 LeuGlnGlyArgPhe- 402
 DB 17828 ACCGTTGGAGGCTTCAGGTTCTAGTCCACGATGTTCCGCGCTATCTAAATCGACCG 17887
 QY 403 GlnGlnAsnAlaGlnAlaPheAspGlnGlnGlyTyrTyrCysSerGlyAspAlaIleLys 422
 DB 17888 GAAAGCACACCGAGGCGTTTCGACGAGAGCGTTGGTACCTTACAGGTATC- 17944
 QY 423 LeuAlaAspProAlaAspProGlnLysGlyLeuMetPheAspGlyArgIleAlaGluAsp 442
 DB 17945 GTGTGTGACAGTGGCGGATGATCGC- 17995
 QY 443 PheLysLeuSerSerGlyValPheValSerValGlyProLeuArgThrArgAlaValLeu 462
 DB 17996 CTAAATCAATTTGGCGCATACCGATCGCGCGAGGAATC- 18049
 QY 463 GlnGlyGlySerTyrValLeuAspVal- 481
 DB 18050 CTAGACATTCGGATGTGGGGGTGGGAGTGTGGCTGCTTAC- 18106
 QY 482 GlyLeuLeuValPheProArgLeuLeuAspCysArgAlaLeuSerGlyLeuGlyLysGlu 501
 DB 18107 GTTCAACGATCTTGCATTCGTC- 18142
 QY 502 AlaSerAspAlaGluValLeuAlaSer- 511

```

; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 2393
; LENGTH: 1902
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-2393

Alignment Scores:
Pred. No.: 5,75e-175      Length: 1902
Score: 1652.50           Matches: 310
Percent Similarity: 70.57%      Conservative: 100
Best Local Similarity: 53.36%      Mismatches: 168
Query Match: 53.62%           Indels: 3
DB: 4                      Gaps: 3

US-09-750-986D-30 (1-589) x US-09-328-352-2393 (1-1902)

QY 5 GUAAlaLeuDeuProPheProGlyArgIleLeuGluLeuGluGluHisThrProAlaLeuThr 24
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133 GAGCAATTAAACCGTATCCACAAAACGACAGATGTTTAATCTTTTGCACAAACC 132

QY 25 ArgProGluGlnThrCysValAlaAlaArgAlaAlaAngIylutPrArgArgIleSer 44
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 193 AAACCCGACCATATTTTTCGACAAAACGCAATGCTCAAGCGAATGGTCAAACTGAGT 252

QY 45 TyrAlaGluMetPheHisValenValArgAlaIleAlaGlnSerLeuLeuProTyrGlyLeu 64
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 253 TAAGCAGAAAGTTTACAAACGCGATGCGACATGCTCAACCTTGCTGCTCGTAATTTA 312

QY 65 SerAlaGluArgProLeuLeuIleValSerGlyAsnAspLeuGluHisLeuGlnLeuAla 84
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 313 AGCCAAAGAAAGACTCTGTCATTTTAAGTGTATATATCTCGAACCTTTTAACACTGTCT 372

QY 85 PheGlyAlaMetTyrAlaGlyIleProTyrCysProValSerProAlaTyrSerLeuLeu 104
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 373 ATGGCTGCGCATGCTGCGACGCGTCTCTCGGCTATTTCCCTGCTACTCTGTGATT 432

QY 105 SerGluAspLeuAlaIleValSerGluHisIleValGlyLeuLeuGlnProGlyLeuValPhe 124
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 433 TCTCAAGACTTGGGCAAACTCAACATGTTGTAAGTGTCTCAACCTGATGATGTTTAT 492

QY 125 AlaAlaAspAlaAlaProPheGlnArgAlaIleGluThrIleLeuProAspAspValPro 144
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 493 GCCAGGATGACAGACTTTTGCACAAAGCATTCAGGCACTGATATACGCTGATATTGAA 552

QY 145 AlaIlePheThrArgGlyLeuAlaGlyArgArgThrValSerPheAspSerLeuLeu 164
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 553 GTGGTACCAATTAAGGAATAGTGGCGATCAGATCTGCACGCTTTTCAATCGCTGTTA 612

QY 165 GluGlnProGlyIleGluAlaAspAsnAlaPheAlaIleThrGlyProAspThrIle 184
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 613 GATACACCAAGTTTCA--AATGTTCAAGATTATTAACCCCTTGTAAGAAACCAAGATT 669

QY 185 AlaIlePheLeuPheThrSerGlySerThrIleLeuProValAlaValProThrThrGln 204
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 670 GCCAAATTTCTGTTTACATCAGGTTCACTAATTTACTTAAGGCTGACGACCAACAT 729

QY 205 ArgMetLeuCysAlaAngGlnIleMetLeuLeuGlnThrPheProValPheGlyGluGlu 224
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 730 TTAATGTTGTTGTTAATCAGCAAAATGTTATTCAGACTTCCCTGAGTTTGAAGAAACA 789

QY 225 ProProValLeuValAspThrLeuProTyrAsnHisIleThrPheGlyGlySerHisIle 244
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 790 CCGCTGCTCTACTCAGCTGCTGCTCTGGCACCAACATTTGGCGGCACTCAAAATGTC 849

QY 245 GlyIleValLeuIleArgGlyGlyThrTyrTyrLeuAspAspGlyLysProThrAlaGln 264
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 850 GGCATGCACTCTATAACGCGGATGATTAATTAATGATATGCAAAACCGTTGCGACGA 909

QY 265 GlyPheAlaGluThrLeuArgAsnLeuSerGluIleSerProThrAlaTyrLeuThrVal 284
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 910 AATTTTGACGAAACTATTCGTAATCTCAAGAAATTTCTCAACTGTTTATTTAAATGTG 969

```

```

QY 285 ProIleGlyTyrGluGluLeuValGlyAlaLeuGluIleArgAspSerThrLeuArgIleArg 304
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 970 CCAAAAGTTGGGAAGAACTCACCGAAGCTTTGAAAAGATGAAAGATTAAAGACCC 1029

QY 305 PhePheAlaArgMetLysLeuPhePheAlaAlaAlaGlyLeuSerGlnGlyIleTyr 324
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1030 TTTTGGCCAAAGTTAAATTTATTTATCTTGGCGGTGCTGACATTTTCAGAAAGCGGCTGG 1089

QY 325 AspArgLeuAspArgValAlaGluGlnHisCysGlyGlyArgIleArgMetMetAlaGly 344
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1090 AACGACTCATTAATTAATGCTCAGCAACATGGCGGAGAAAATAATCCGCATTATGAGCGGA 1149

QY 345 LeuGlyMetThrGluThrAlaProSerCysThrPheThrThrGlyProLeuSerMetAla 364
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1150 TTGGCATGACCAAACTGCTCCCTTGTGCTTTTACACTGACCCACCGGTAATGGCT 1209

QY 365 GlyTyrIleGlyLeuProAlaProGlyCysGlyValIleLeuValProValAspGlyLys 384
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1210 GCGTTATTTGTTACCTGCTCCGGATGGGAATTAAGTATGTTCCATGTGTTGACAAA 1269

QY 385 LeuGluGlyArgPheHisGlyProHisValMetSerGlyTyrTyrArg--AlaProGlu 403
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1270 CTGAGATTTTGGCTTCGTGCAAAACAGTCATGAAAGGCTATTGGCGCTTAAAGCGGAC 1329

QY 404 GluAsnAlaGlnAlaPheAspGluGluGlyTyrTyrCysSerGlyAspAlaIleLysLeu 423
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1330 CAACAAAGTACTATTTTGTATGATGAAGGCTTTTCATACAGGCGATGCCGTTCTTTA 1389

QY 424 AlaAspProAlaAspProGlnIleGlyLeuMetPheAspGlyArgIleAlaGluAspPhe 443
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1390 GTGATGTCATGATCATCAACCAAGGCTTAATGACGACGAATATGCGGAAGACTTT 1449

QY 444 LysLeuSerSerGlyValPheValSerValGlyProLeuArgThrArgAlaValLeuGlu 463
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1450 AAACCTAATTAACAGCACTTTTGTGAATGTGCGACACTACGCAAAAGGCTTATTTCAA 1509

QY 464 GlyIleSerTyrValLeuAspValValAlaAlaProAspArgGluCysLeuGlyLeu 483
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1510 GGTAAATTTACTCATCAAGATGTTTGTATTTACTGTTCAAACTGAATGCTATTTGTTT 1569

QY 484 LeuValPheProArgLeuLeuAspCysArgAlaLeuSerGlyLeuGly--LysGluAla 502
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1570 CTGATTTTTCAAATTAATGACCGTTGGCTCAATATGACAGGCTTAAAGTGGCGAATAT 1629

QY 503 SerAspAlaGluValLeuAlaSerGluProValArgAlaThrPheAlaAspThrLeuLys 522
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1630 TCTGACGACGAGTACGACCAACCTTAAGTCAACATGTTCCGCCAATTTTAAACG 1689

QY 523 ArgLeuAspArgGluAlaThrGlyAsnAlaSerArgIleMetThrValGlyLeuLeuAsp 542
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1690 ACCTTAATTAAGATGCGACTGCGACGCTCAATACAGTCTCAATGCTTTATTAATGACC 1749

QY 543 ThrProSerIleAspLysGlyGluValThrAspLysGlySerIleAsnGlnArgAla 562
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1750 GAGCAACCTCAGTAAATGCGCGGCGAAGTGAACGATTAAGGCAACCTCAATCAAGACAGT 1809

QY 563 ValLeuGlnThrArgSerAlaLysValAspAlaLeuTyrArgGlyGluAspGlnSerMet 582
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1810 ATTAACCAAGGTCGTCGCTTTAATTAATGATGACCTTTATCAAAAGCACTGATTAACCG 1869

QY 583 Leu 583
   |||
Db 1870 CTG 1872

```

```

RESULT 4
US-08-311-731A-24
; Sequence 24, Application US/08311731A
; Patent No. 6583266
; GENERAL INFORMATION:
; APPLICANT: SMITH, DOUGLAS
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR

```

NAME/KEY: CDS
LOCATION: (3146) .. (3997)
OTHER INFORMATION: Gene = "ORF1"
US-08-976-063E-1

Alignment Scores:

Pred. No.:	0	Length:	32679
Score:	3082.00	Matches:	589
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	4	Gaps:	0

US-09-750-986D-30 (1-589) x US-08-976-063E-1 (1-32679)

```
QY 1 MetArgSerLeuGluAlaLeuLeuProPheProGlyArgIleLeuGluArgLeuGluHis 20
Db 21286 ATGGCTTCTCTCGAGGCGCTTCTTCCCTTCCCGGGTCGAATTTCTTGAGCGTCTGAGCAT 21347
QY 21 TrpAlaIysThrArgProGluGluGlnThrCysValAlaAlaArgAlaAlaAsnGlyGluTrp 40
Db 21348 TGGGCTAAGACCCGTCAGAAACAACCTGCGTGTCTGCAAGGGGGCAAAATGGGGAATGG 21407
QY 41 ArgArgIleSerTrpAlaGluMetPheHisAsnValArgAlaAlaIleAlaGlnSerLeuLeu 60
Db 21408 CGTGTATCAGCTACGGGGAATGTTCCACAACGTCGCGCATCGCACAGAGCTTGCTT 21467
QY 61 ProTrpGlyLeuSerAlaGluArgProLeuLeuIleValSerGlyAsnAspLeuGluHis 80
Db 21468 CTTTACGAGCTATCGGAGACCGTCGCTGCTTACGCTCTGGAATAAGACCTGGAACAT 21527
QY 81 LeuGlnLeuAlaPheGlyAlaMetTrpAlaGlyIleProTrpCysProValSerProAla 100
Db 21528 CTTACGCTGGCATTTGGGGCTATGATAGCGGCAATTCCTTATCCCGGTCTCTCGTCT 21587
QY 101 TyrSerLeuLeuSerGlnAspLeuAlaIleLeuArgHisIleValGlyLeuLeuGlnPro 120
Db 21588 TATTTCATGCTGTCGCAAGATTTGGCGAAGCTGCTACATCGTAGGTCTTCTGCAACCG 21647
QY 121 GlyLeuValPheAlaAlaAspAlaAlaProPheGlnArgAlaIleGluThrIleLeuPro 140
Db 21648 GGAAGCTCTTGTGCTGCAGATGACGACCTTTCCAGCGCGCAATTTGAGACCATTTGCGG 21707
QY 141 AspAspValProAlaIlePheThrArgGlyGluLeuAlaGlyArgArgThrValSerPhe 160
Db 21708 GAGCAGCTGCCCGCAATCTTCACTCGAGGGCAATTGGCCGGGGCGCGACGCTAAGTTT 21767
QY 161 AspSerLeuLeuGluGlnProGlyGlyIleGluAlaAspAsnAlaPheAlaAlaThrGly 180
Db 21768 GACAGCTGCTGGAGCAGCCTGGTGGGATTGAGGCAATATGCTTTGCGGCAACTGGC 21827
QY 181 ProAspThrIleAlaIleAspPheLeuPheThrSerGlySerThrIleLeuProValAla 200
Db 21828 CCGGATTCGATTTGCCAAGTCTTGTCACCTTGCTTACCAAACTGCTTAAGCGGATG 21887
QY 201 ProThrThrGlnArgMetLeuCysAlaAsnGlnIleMetLeuLeuGlnThrPheProVal 220
Db 21888 CCGAGTACTCAGGGAATGCTCTGCGCAATCAGCAATGCTTTCGAAACTTTCCGGTT 21947
QY 221 PheGlyGluGluProProValLeuValAspTrpLeuProTrpAsnHisThrPheGlyGly 240
Db 21948 TTTTGTTGAAAGCGCGCGTGTGCTGTGAGCTGTTGCGTGAACCACTTCCGCGCG 22007
QY 241 SerHisAsnIleGlyIleValIleLeuTrpAsnGlyGlyIleThrTrpTrpLeuAspAspGlyLys 260
Db 22008 AGCCACAACATCGGCATCGTGTGTATACAACGCGCGCACTGTAATCTTGCAGAGGTAAA 22067
QY 261 ProThrAlaGlnGlyPheAlaGluThrLeuArgAsnLeuSerGluIleSerProThrAla 280
Db 22068 CCAACGCCCAAGGTTCCGCGAGACGCTTCGCAACTTGAGCCAAATCTCTCCACACTGCG 22127
QY 281 TyrLeuThrValProLysGlyTrpGluLeuValGlyAlaLeuGluArgAspSerThr 300
|||||
```

```
Db 22128 TACCTCACTGTGCCGAAGCGCTGGAGGAATTAAGTGGTCCCTTGAGCGAGACGATACC 22187
QY 301 LeuArgGluArgPhePheAlaArgMetLysLeuPhePhePheAlaAlaAlaGlyLeuSer 320
Db 22188 CTGCGCAACGCTTCTTCCCTCGCATGAAGCTGTTCTTCTTCCGCGCGGCTGGGTTCG 22247
QY 321 GlnGlyIleTrpAspArgLeuAspArgValAlaGluGlnHisCysGlyGlyLysArgIleArg 340
Db 22248 CAAGGATCTCGGATCGTTTGACCGGGTGTCTGAACGACACTGTGGTGAACGATTCGC 22307
QY 341 MetMetAlaGlyLeuGlyMetThrGluThrAlaProSerCysThrPheThrThrGlyPro 360
Db 22308 ATGATGGCGGGTCTGGGATACAGACAGCTGCTCTTCGCACTTTTACCAACCGGACCG 22367
QY 361 LeuSerMetAlaGlyTrpIleGlyLeuProAlaProGlyCysGluValLysLeuValPro 380
Db 22368 CTGTGATGCTGTGTTACATTGGGCTGCCAGCGCTGCTGCTGAGTCAACTGCTTCCG 22427
QY 381 ValAspGlyLysLeuGluGlyArgPheHisGlyProHisValMetSerGlyTrpArg 400
Db 22428 GTCCATGGGAAATTGGAAGGGCGTTTCATGTCGCGACGTCATGAGCGGCTACTGGCGT 22487
QY 401 AlaProGluGlnAsnAlaGlnAlaPheAspGluGluGlyTrpTrpCysSerGlyAspAla 420
Db 22488 GCTCCTTAAACAAATGCCCCAAGCGTTGACAGAGGAAGCTATTAATGCTCGGTGATGCC 22547
QY 421 IleLysLeuAlaAspProAlaAspProGluLysGlyLeuMetPheAspGlyArgIleAla 440
Db 22548 ATCAAAATGGCAGATCTTCCGATCTTCAGAAAGGTCTGATGTTGACGTCGAAATGCT 22607
QY 441 GluAspPheLysLeuSerSerGlyValPheValSerValGlyProLeuArgThrArgAla 460
Db 22608 GAAAGCTTTCAAGCTGTCCTCAGGGGCTATTGTCTCAGCGTTGGGCCATTGCCGACCGGCG 22667
QY 461 ValLeuGluGlyGlySerTrpValLeuAspValValAlaAlaProAspArgGluCys 480
Db 22668 GTTCTGGAAGCGGCTTTCAGTCTGACGTAGTGTGCTGCTCTGATCGGAATGTC 22727
QY 481 LeuGlyLeuLeuValPheProArgLeuLeuAspCysArgAlaLeuSerGlyLeuGlyLys 500
Db 22728 CTTGATTTGCTCGTGTTCGCGCTTCTTCTGACATGCGCTGCTTGTGGGGCTAGGAAAA 22787
QY 501 GluAlaSerAspAlaGluValLeuAlaSerGluProValArgAlaTrpPheAlaAspTrp 520
Db 22788 GAGGCGTCGAGCGCGCGAGGTGCTTGCAGTGAAGCGGTTGGGCTGGTTTGCAGACTGG 22847
QY 521 LeuLysArgLeuAsnArgGluAlaThrGlyAsnAlaSerArgIleMetTrpValGlyLeu 540
Db 22848 CTCMAAGACTCAATCGAAGAAGCAACTGGCAATGCCAGTGCATCATGTGGTGAAGCTC 22907
QY 541 LeuAspThrProProSerIleAspLysGlyGluValThrAspLysGlySerIleAsnGln 560
Db 22908 CTCGATTCGCGCGGCTGATTTGATGAAGGCGAGGTCACTGACAAAGGCTCGATTAACAG 22967
QY 561 ArgAlaValLeuGlnTrpArgSerAlaLysValAspAlaLeuTrpArgGlyGluAspGln 580
Db 22968 CGGCTGTTTTCGAATGCGGTCGGAAGGTGATGCGCTGATCTGTGTGAAGATCAA 23027
QY 581 SerMetLeuArgAspGluAlaThrLeu 589
Db 23028 TCCATGCTGCTGTGACGAGGCCACACTG 23054
```

RESULT 3

US-09-328-352-2393
Sequence 2393, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04

US-09-750-986D-30 (1-589) x US-08-976-063E-29 (1-1770)

QY 1 MetArgSerLeuGluAlaLeuLeuProPheProGlyValArgIleLeuGluValGlyLeuHis 20
 Db 1 ATCGCTTCTTCAGAGCGCTTCTTCCCTCCGGGTGAAATTCCTGAGCGCTTCGAGCAT 60
 QY 21 TPAlaLysThrArgProGluGluThrCysValAlaAlaArgAlaAlaAsnGlyGluThr 40
 Db 61 TGGGCTAAGACCCGTCAGAACAAACCTGCTGTGCGCAGGCGGCAAAATGGGAAATGG 120
 QY 41 ArgArgIleSerTyraIaGluMetPheHisAsnValArgAlaIleAlaGlnSerLeuLeu 60
 Db 121 CGTCGATACGCTACGCGGAAATGTTCCAAACGTCGCCGCCCATCGACAGAGCTTCTT 180
 QY 61 ProTyGlyLeuSerAlaGluArgProLeuLeuIleValSerGlyAsnAspLeuGluHis 80
 Db 181 CCTTACGGAATATCGGACAGCGCTCCGCTTATCGTCTGTGAAATGACCTGGAACAT 240
 QY 81 LeuGluLeuAlaPheGlyAlaMetTyraIaGlyIleProTyraCysProValSerProAla 100
 Db 241 CTTCAGCTGACATTTGGGCTATGATGCGGCAATTCCTATGCGCGGTCTCTGCT 300
 QY 101 TyrSerLeuLeuSerGlnAspLeuAlaValLeuArgHisIleValGlyLeuLeuGlnPro 120
 Db 301 TATTCACTGCTGCGCAAGATTTGGCGAAGCTGCTCACAATCGTAGGCTTCTGCAACCG 360
 QY 121 GlyLeuValPheAlaAlaAspAlaAlaProPheGlnArgAlaIleGluThrIleLeuPro 140
 Db 361 GGACTGCTCTTCTGCTCGCATGACAGCACTTTCACAGCGCGCAATTGAGACATTTGCGC 420
 QY 141 AspAspValProAlaIlePheThrArgGlyGluLeuAlaGlyArgAspGluThrValSerPhe 160
 Db 421 GACGAGCGTCCCGCAATCTTCACTCGAGGCAATGCGCGCGCGCGCACAGGTGAGTTT 480
 QY 161 AspSerLeuLeuGluGlnProGlyGlyIleGluAlaAspAsnAlaPheAlaAlaThrGly 180
 Db 481 GAAAGCTGCTGGAGACGCTGCTGGGATTTGAGGCGAATATGCTTTGGCGCAACTGGC 540
 QY 181 ProAspThrIleAlaLysPheLeuPheThrSerGlySerThrLysLeuProLysValAla 200
 Db 541 CCGGATACGATTCGCAAGTCTTGTTCATCTTGGCTTACCAAACTGCTTAAGCGCGTG 600
 QY 201 ProThrThgIaArgMetLeuCysAlaAsnGlnGluMetLeuGlnThrPheProVal 220
 Db 601 CCGACTACTGAGGAATGCTCTGCGCAATCAGCAAGATGCTTCTGCAAACTTTCGCGGT 660
 QY 221 PheGlyGluGluProProValLeuValAspTrpLeuProTrpAsnHisThrPheGly 240
 Db 661 TTGTGGAAAGCGCGCGGTGCTGGTGGACTGGTTCGCGGAACCACTTCGCGCGC 720
 QY 241 SerHisAsnIleGlyIleValLeuTyraAsnGlyGlyThrTyraLysAspAspGlyLys 260
 Db 721 AGCCACAACATCGGCATCGTGTGTACAAACGCGCGCACTTGAACCTTGAACGCGTAAA 780
 QY 261 ProThrAlaGlnGlyPheAlaGluThrLeuArgAsnLeuSerGluIleSerProThrAla 280
 Db 781 CCAACGCGCCCAAGGTTGCGCGAAGCCTTGGCACTTGACGCAAACTCTCCACACTGCG 840
 QY 281 TyrLeuThrValProLysGlyTyraGluGluValGlyAlaLeuGluArgAspSerThr 300
 Db 841 TACTCACTGCTGCGAAAGCTGCGGAGGATTTAGTGGTCCCTTGAAGCAACAGTACC 900
 QY 301 LeuArgGluArgPhePheAlaArgMetLysLeuPhePhePheAlaAlaAlaGlyLeuSer 320
 Db 901 CTGCGGGAAGGCTTCTTCTGCTCCATGAACCTGTTCTTCTCCGCGGCGTGGGTGTCTG 960
 QY 321 GlnGlyIleTrpAspArgLeuAspArgValAlaGluGlnHisCysGlyGlyValArg 340
 Db 961 CAAAGGATCTGGGATCTTTGACCGGGTCTGGAACAGCACTGTGTGTAGCGCATTCGC 1020
 QY 341 MetMetAlaGlyLeuGlyMetThrGluThrAlaProSerCysThrPheThrThrGlyPro 360

Db 1021 ATGATGCGCGGTCTGGGCAATGACGAGAGACTGCTCTTCTCGACATTTTACACCGGACCG 1080
 QY 361 LeuSerMetAlaGlyTyraIleGlyLeuProAlaProGlyCysGluValLysLeuValPro 380
 Db 1081 CTGTGCATGGCTGGTTACATTTGGGCTGCGACGCCCTGGCTGCGCAAGGTCAAGCTCGTCCG 1140
 QY 381 ValAspGlyLysLeuGluGlyArgPheHisGlyProHisValLysSerGlyTyraTrpArg 400
 Db 1141 GTTCATGGAAATTTGAAAGGCGTTTCCATGTCGCCGCACTCATGAGCGGCTACTGGCCT 1200
 QY 401 AlaProGluGlnAlaAlaAlaPheAspGluGluGlyTyraTyraCysSerGlyAspAla 420
 Db 1201 GCTCTGAACAAATGCCCAAGCGTTTGCAGAGGAAGCTTATTAAGCTCTCGGTGATGCC 1260
 QY 421 IleLysLeuAlaAspProAlaAspProGluLysGlyLeuMetPheAspGlyValGlnAla 440
 Db 1261 ATCAAAATGGCAATCTTCCGATCTCTCAAGAAAGTCTGATGTTTACCGGTCAAAATGCT 1320
 QY 441 GluAspPheLysLeuSerSerGlyValPheValSerValGlyProLeuArgTrpArgAla 460
 Db 1321 GAAAGCTTCAAGCTTCTTCCAGGGGTATTTGTCAAGCTTGGGCACTTGGCAGCGGCGC 1380
 QY 461 ValLeuGluGlyGlySerTyraIleAspValValAlaAlaAlaProAspArgGluCys 480
 Db 1381 GTTCTGAAAGCGGCTTTAACGCTTGAAGCTGAGAGTGTGCTGCTCTGATCGTGAATGC 1440
 QY 481 LeuGlyLeuLeuValPheProAlaGluLeuAspCysArgAlaLeuSerGlyLeuGlyLys 500
 Db 1441 CTTGGAATGCTGCTGTTTCCGCTCTTCTGACTGCGGCTGTGCGGGCTAGAAATA 1500
 QY 501 GluAlaSerAspAlaGluValIleAlaSerGluProValArgAlaTrpPheAlaAspTrp 520
 Db 1501 GAGGCGTGGAGCCCGCAAGGCTTCCCACTGAGGACCGGCTGGGCTGTGCTGACTGG 1560
 QY 521 LeuLysArgLeuAsnArgGluAlaThrGlyAsnAlaSerArgIleMetTrpValGlyLeu 540
 Db 1561 CTCAAACGACTCATCGAAGAGCACTGGCAAGCCAGTGCATCATGTGGTGAAGGCTC 1620
 QY 541 LeuAspThrProProSerIleAspLysGlyGluValThrAspLysGlySerIleAsnGln 560
 Db 1621 CTGGAATACCGCGCGGTGATGATTAAGGCGAGGCTCACTGCAAGGAGGCTGATCAACAG 1680
 QY 561 ArgAlaValLeuGlnTrpArgSerAlaLysValAspAlaLeuTyraArgGlyLysArgGln 580
 Db 1681 CGGCGCTGTTTGCATAGCGCGGTGCGGCAAGTGAATGATGCTGTGTGAAGTCAA 1740
 QY 581 SerMetLeuArgAspGluAlaThrLeu 589
 Db 1741 TTCATGCTGCTGACAGGCGCACACTG 1767

RESULT 2
 US-08-976-063E-1
 Sequence 1, Application US/08976063E
 Patent No. 6524831
 GENERAL INFORMATION:
 APPLICANT: Pfeleferc, Horst
 APPLICANT: Rademorenc, Jurgen
 TITLE OF INVENTION: SYNTHETIC ENZYMES FOR THE PRODUCTION OF CONIFERYL
 TITLE OF INVENTION: ALCOHOL, CONIFERYLALDEHYDE, FERULIC ACID, VANILLIN AND
 FILE REFERENCE: Bayer-9998-CAO
 CURRENT APPLICATION NUMBER: US/08/976, 063E
 CURRENT FILING DATE: 1997-11-21
 PRIOR APPLICATION NUMBER: 196 49 655.1 GERMANY
 PRIOR FILING DATE: 1996-11-29
 NUMBER OF SEQ ID NOS: 45
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 1
 LENGTH: 32679
 TYPE: DNA
 ORGANISM: Pseudomonas sp.
 FEATURE:

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 7, 2005, 23:57:56 ; Search time 142 Seconds
(without alignments)
2948.272 Million cell updates/sec

Title: US-09-750-986d-30
Perfect score: 3082
Sequence: 1 MSLEALLPFGRIERLEH.....KVDALYRGEDQSLRDEATL 589

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=x1p
-O=/cg2_1/USPTO.spool_p/US09750986/runat.03022005.071744.18593/app.query.fasta.1.775
-DB=Issued Patents NA -OPMT=faaetap -SUFIX=tni -MIMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTPMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09750986 @CGN 1.1 105 @runat.03022005.071744.18593 -NCPU=6 -ICPU=3
-NO MMAP -LARGEIOBURY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGIOG
-DEV TIMEOUT=120 -MARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:
1: /cg2_6/prodata/1/ina/5A.COMB.seq:*
2: /cg2_6/prodata/1/ina/5B.COMB.seq:*
3: /cg2_6/prodata/1/ina/6A.COMB.seq:*
4: /cg2_6/prodata/1/ina/6B.COMB.seq:*
5: /cg2_6/prodata/1/ina/PCTUS.COMB.seq:*
6: /cg2_6/prodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	3082	100.0	1770	4	US-08-976-063E-29 Sequence 29, Appl
2	3082	100.0	32679	4	US-08-976-063E-1 Sequence 1, Appl
3	1652.5	53.6	1902	4	US-09-328-352-2393 Sequence 2393, Ap
4	302.5	10.0	38494	4	US-08-311-731A-24 Sequence 24, Appl
5	302.5	9.8	35961	4	US-08-311-731A-138 Sequence 138, Appl
6	283	9.2	1830121	4	US-09-557-884-1 Sequence 1, Appl
7	283	9.2	1830121	4	US-09-643-990A-1 Sequence 1, Appl
8	283	9.2	1830121	4	US-10-329-960-1 Sequence 1, Appl
9	266	8.6	1845	4	US-09-543-681A-2894 Sequence 2894, Ap
10	263.5	8.5	1639	4	US-09-396-154-10 Sequence 10, Appl
11	263.5	8.5	1639	4	US-09-396-154-43 Sequence 43, Appl
12	262.5	8.5	1639	4	US-09-396-154-11 Sequence 11, Appl

13	259	8.4	4403765	3	US-09-103-840A-2 Sequence 2, Appl
14	259	8.4	4411529	3	US-09-103-840A-1 Sequence 1, Appl
15	258	8.4	1639	4	US-09-396-154-42 Sequence 42, Appl
16	252.5	8.2	1639	4	US-09-396-154-6 Sequence 6, Appl
17	251	8.1	2416	4	US-08-592-126-61 Sequence 61, Appl
18	251	8.1	2416	4	US-09-168-595-61 Sequence 61, Appl
19	249.5	8.1	1639	4	US-09-396-154-8 Sequence 8, Appl
20	249.5	8.1	1666	4	US-09-602-628-1 Sequence 1, Appl
21	248.5	8.1	1639	4	US-09-396-154-7 Sequence 7, Appl
22	247.5	8.0	1639	4	US-09-396-154-9 Sequence 9, Appl
23	247.5	8.0	1666	4	US-09-602-628-3 Sequence 3, Appl
24	246.5	8.0	1650	4	US-08-354-240A-3 Sequence 3, Appl
25	246.5	8.0	1653	1	US-08-354-240A-5 Sequence 5, Appl
26	246.5	8.0	1685	4	US-09-602-628-5 Sequence 5, Appl
27	246.5	8.0	1686	4	US-09-602-628-7 Sequence 7, Appl
28	246.5	8.0	2337	4	US-10-066-130-20 Sequence 20, Appl
29	246.5	8.0	2674	4	US-10-066-130-19 Sequence 19, Appl
30	246.5	8.0	2771	4	US-10-066-130-18 Sequence 18, Appl
31	246.5	8.0	5860	4	US-10-066-130-17 Sequence 17, Appl
32	246.5	8.0	12614	4	US-09-577-424-1 Sequence 1, Appl
33	246.5	8.0	12844	4	US-09-743-194-19 Sequence 19, Appl
34	246.5	8.0	12850	4	US-09-743-194-22 Sequence 22, Appl
35	246.5	8.0	12851	4	US-09-743-194-21 Sequence 21, Appl
36	246.5	8.0	13073	4	US-09-743-194-20 Sequence 20, Appl
37	246.5	8.0	13654	4	US-09-743-194-28 Sequence 28, Appl
38	246.5	8.0	14194	4	US-09-577-424-3 Sequence 3, Appl
39	244.5	7.9	2782	4	US-10-140-002-205 Sequence 205, App
40	244	7.9	3634	4	US-09-814-915A-28 Sequence 28, Appl
41	244	7.9	8937	4	US-09-872-733A-8 Sequence 8, Appl
42	244	7.9	8937	4	US-09-872-733A-9 Sequence 9, Appl
43	243.5	7.9	2445	1	US-08-122-520C-8 Sequence 8, Appl
44	241.5	7.8	1722	3	US-08-718-425-1 Sequence 1, Appl
45	240.5	7.8	1650	1	US-08-354-240A-1 Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-976-063E-29
; Sequence 29, Application US/08976063E
; Patent No. 6524831
; GENERAL INFORMATION:
; APPLICANT: Steinduchel, Alexander
; APPLICANT: Priefert, Horst
; TITLE OF INVENTION: SYNTHETIC ENZYMES FOR THE PRODUCTION OF CONFERYL
; TITLE OF INVENTION: ALCOHOL, CONFERYLALDEHYDE, FERULIC ACID, VANILLIN AND
; FILE REFERENCE: BAYER-9998-CAO
; CURRENT APPLICATION NUMBER: US/08/976,063E
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: 196 49 655.1 GERMANY
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 29
; LENGTH: 1770
; TYPE: DNA
; ORGANISM: not required under old rule
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1767)
; OTHER INFORMATION: product = "Ferulaseure-CoA-Synthetase" / gene =
; OTHER INFORMATION: "fcs"

US-08-976-063E-29
Alignment Scores:
Pred. No.: 0
Score: 3082.00 Length: 1770
Percent Similarity: 100.00% Matches: 589
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0

This Page Blank (uspto)

Db		904	-----TCAGATGTTTACCATCATATCTTCCCTTTGGCCACATATATAAGCT	951
Oy		241	SerHisabaNIIEgLYIleValLeuTYraNgLyglYThrTYrLeuaSpaspLyIs	260
Db		952	GCAATATCAGATT--ATGGGGGTGATGATGGT	981
Oy		261	ProThrIaagInglyPhe-----AlaGIuThnLeuaRgaSnuSerLyu	275
Db		982	--GTTGCTGTCGGGTTTCTATTCAGGGGATGCTTTGAAGCGTAGTGATGAAATTTTGTCTGTG	1038
Oy		276	IleSerProThrIaTyrlEuThrValProlysGIYTPGluGlueValIGlYAlaLeu	295
Db		1039	TTAAGACCACAATTTCTGTAGTGTCCCTCGCTTRTATTCGAATATATAGTGCATT	1098
Oy		296	GIuArGaSpSerThrLeuaRgGLuArGPhePheAlaRmeLysLeuPhePheAla	315
Db		1099	-----ACAAGTGCCGTAAAATCATCTGGGGTTGTGAAAAAAGGCTTTTGCAAATTTGCC	1152
Oy		316	AlaIaagLYLeuSerGlnLYIle-----TryAspaRg	326
Db		1153	TATAACTCAAAGAACAAGCACCATCTTAATGGCGGAGCTCCTTCTGCATTTTGGACAAG	1212
Oy		327	Leu-----AsPaRgVaIaIgLIuGlnHiasCySGlyGuArGIeArgMetMetala---	343
Db		1213	CTGGTCTTCACAAAATTAAGAAAAAG---CTTGSTGACCGGTTGGTTTATGGGCTCT	1268
Oy		343	-----	343
Db		1270	GGTGCCTTCTCCTTGTCTACCTGATGTCAAGATTCTTGAGAATATGCTTTGAGATGTTCCG	1329
Oy		344	-----GlyLeuGIYmetThrGIuThrAlaProSerCYsthrPheThr----	358
Db		1330	GTGCGTGAAGGGATGTATGATACCGAGACT----TCTTGTGTCAATGAAGGCTATGAT	1383
Oy		359	---GlyProLeuSerMetAlaGIYTrtIegLYLeuProLaProGLYCysGIUValIys	377
Db		1384	GATGGTGC---AATTATCTGGCCACGCTCGATTCCTTAATCCAAGCTTGGAGGTAAAG	1440
Oy		378	LeuValProValaSp-----GlyLys	384
Db		1441	CTTGTGATGTTCCTCCGAAATGAATTAACATACGGAAGATCAACATACCACGTTGTGTA	1500
Oy		385	LeuGIuGLYArGPheHieGLYProHISValMetSerGIYTrTPRaGAlaPProGLuIn	404
Db		1501	ATCTGCTGAACA-----GGACCAATCATCTTCAAAAGGCTATYTCAAAAGATGAAGAACAA	1554
Oy		405	AsnaIaGlnAlaPheaSPGLuGIuGLYTrTYrCYSSerGIYAAPAlaIleLysLeuAla	424
Db		1555	ACGAGAGAAATTTCTTGATGAGATGGCTGGCTACACAGAGATATC-----	1602
Oy		425	AspProLaSpProGLINysGIYLeuMeRheaSpGIY-----	437
Db		1603	-----GGGTGTGGTTACTCGTGGTGTGGCTCAAGATCATAGAC	1641
Oy		438	ArgIlealaglueApPheLysLeuSerSerGIYAlaPheValSerValGIYProLeuArg	457
Db		1642	AGGAAGAAGAACATATTTAAAGTTGGCGMAAGAAATATATAGCACAGAGAAATGCAGAA	1701
Oy		458	-----ThraRgaIaValLeuGIuGLYglYSerTYr-----	467
Db		1702	AATGTTTATACCAATGTAGATTGCTTTGGAGGTTTCATTCACGGGATGATGCTTCAT	1761
Oy		468	-----ValLeuApRvaIaValaIaAlaProaSp-----	477
Db		1762	TCTCTCTAGTAGATATATGTTTCACTAGTGCAGCCCGAAGTATGAAGAAAGATGGGCTGCATCA	1821
Oy		478	-----ArgGIuCYsLeuGIYLeuLeuVal---PheProArGLeuLeuAap--	491
Db		1822	GAAGGCATCAAGTATGAGCATCTAGACACGCTCTGTAAAGATCCAAAGAGTGCAGAAAGACT	1881
Oy		492	---CysARgaIaLeuSerGIYLeuGIYLYsgLuIa-----Ser	503

```

Db      1882 GTTCTTGCTGATGATGATGACCTTGAAGAAGAACTCAGTTGAGAGGGTTTGAGTTGCA 1941
Oy      504 Asph1aGluValLeuAlaSerGluPro----- 512
          |||  |||  |||  |||  |||
Db      1942 AAGCTGTGACTTTGGTGCCAGAACCAATTCACTTGGAGATGAGCTTTCACCAACA 2001
Oy      513 -----ValArgAlaTTrpPheAlaAspTrpLeuIlyArgLeuAsn 525
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2002 TTCAAGATTAAGAGACCTCAAGCAAAAGCCTACTTGCAGAGCAATTAAGCAAAATGTAT 2061
Oy      526 Arg1Iu 527
          |||
Db      2062 GCGGAA 2067

```

```
Search completed: February 8, 2005, 00:16:11
Job time : 977 secs
```

```
Db      1882 GTTCTTGCGTAGATGAGTAACCTTGGAGAAGAACTCAGTTGAGAGGGSTTTGAGTTTCGA   1941  
OY      504 AspaIagluValLeuaIasergluPro----- 512  
        |||         |||         |||||  
Db     1942 AAGCGGTGACTTTGGTGCCAGAACCAATTGCACCTTGAGAATGGACTTCTCACACAACA   2001  
OY      513 -----ValArgAlaTriPhealAspTrpleuLysArgleuasn 525  
        ::::||:::||:::||:::||::|||:::|||||:  
Db     2002 TTCAAGATAAAGAGACCCTCAAGCAMAAAGCCTACTTTGCCAGAGCCAATTAGCAMAATGTAT 2061  
OY      526 Argglu 527  
        |||  
Db     2062 GCGGAA 2067
```

Search completed: February 8, 2005, 00:16:11
Job time : 977 secs

```

Db      1873 ATGGTATGAAAGCGTTATTAACAAGAGCCAGAAAGAAACGGCTCAAGCTTACCGAAGAT 1932
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Qy      413 G|Y|T|Y|C|y|s|e|r|g|y|a|p|a|l|a|l|e|y|l|e|u|a|a|p|p|o|l|a|a|p|p|o|l|y|g|y 432
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      1933 GGTTTCTTAAAACTGGCGATCGACGAGAAATTT-----GACGAAACAAAGCAAT 1980
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Qy      433 l|e|u|e|r|p|h|e|a|s|e|r|g|y|l|e|a|g|l|u|a|p|h|e|l|e|u|s|e|r|g|y|a|l|p|h|e|a|s|e|r 452
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      1981 TTAATTATTAACCGATGCTATCAAGAAATTAATGAAAACTCAAGCGCAAAATATATCGCA 2040
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Qy      453 v|a|l|g|y|p|r|o|l|e|u|r|g|h|r|a|r|g|a|l|a|v|a|l|e|u|g|l|y|g|y|s|e|r|y|a|l|e|u|a|p|a|l|a| 472
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      2041 -----CCACATATATTCGAAAGCAAAATCGTAAAGATTAATTTACGAAACAAATTGCG 2094
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Qy      473 v|a|l|a|l|a|l|p|r|o|a|p|h|a|r|g|l|y|l|e|u|g|y|l|e|u|a|l|p|h|e|p|r|o|l|g|l|e|u|a|p|y|e 492
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      2095 ATCATCGCTGATCGCAAAAAATATGTATCCGCGCTTATGTGCT-----TGC 2142
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Qy      493 ---A|r|g|a|l|e|u|s|e|r|g|y|l|e|u|g|y|g|l|u|a|l|s|e|r-----A|p|a|l|a|l|u 506
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      2143 TTGTATAGTTTGGAGAAATACGCTTAAACAGCTCAATATTAAATATCATGACCGTTAGAA 2202
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Qy      507 v|a|l|e|u|a|l|s|e|r|g|l|u|p|r|o|v|a|l|a|r|g|a|l|a|t|r|p|h|e|a|l|a|p|r|l|e|u|y|s|a|r|g|l|e|u|a|a|n|a|r|g 526
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      2203 C|T|A|C|T|A|A|A|A|A|T|T|C|A|C|A|T|T|C|G|A|A|A|A|T|T|G|A|C|A|T|G|T|A|T|A|T|G|C|G|C|A|A|A|A 2262
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Qy      527 g|l|u|a|l|a|t|h|r|g|y|a|n|a|l|s|e|r|a|r|g|l|e|u|e|r|t|r|p|v|a|l|g|l|e|u|a|p|h|t|r|p|r|o|s|e|r 546
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      2263 GAA---T|T|G|C|T|C|A|T|T|C|G|A|G|A|G|T|A|A|A|A|A|A|T|T|C|A|G|T|A|C|T|T|C|T|C|A|A|G|A|T|T|C|A|G 2319
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Qy      547 l|l|e|a|p|l|y|g|l|y|l|u|a|l|t|h|r|a|p|l|y|s|g|l|s|e|r|l|e|a|n|g|l|u|r|g|a|l|a|l|e|u|g|l|n|t|r|p 566
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      2320 A|T|T|A|A|A|T|T|A|G|C|G|A|A|A|T|T|A|C|A|C|A|C|A|T|T|A|A|A|T|T|A|G|T|A|A|A|A|G|T|A|T|T|A|G|A|A|C|G|T 2379
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Qy      567 A|r|g|s|e|r|a|l|a|y|a|l|a|p|a|l|e|u|t|y|r|a|r|g|l|y|l|u|a|p 579
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      2380 T|A|T|C|G|A|G|A|C|A|A|T|T|G|A|G|C|A|A|T|C|A|T|C|A|T|T|C|A|C|A|G|A|A 2418
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

```

PT      a plant, making a transgenic plant and for producing variants of acyl-CoA
PT      synthetases.
PS      Claim 1; SEQ ID NO 6; 155bp; English.
XX      CC
XX      The invention comprises the amino acid and coding sequences of plant acyl
XX      Co-coenzyme A synthetase (acyl-CoA synthetase) enzymes. The DNA and protein
XX      sequences of the invention are useful for altering a phenotype of a plant
XX      (transgenic plant). The DNA and protein sequences of the invention are
XX      also useful for producing variants of acyl-CoA synthetases. The present
XX      CC DNA sequence encodes a plant acyl-CoA synthetase of the invention.
SQ      Sequence 2103 BP; 588 A; 417 C; 499 G; 597 T; 0 U; 2 Other;

Alignment Scores:
Pred. No.:      3,43e-15      Length:      2103
Score:          279.00      Matches:      150
Percent Similarity: 37.76%      Conservative: 100
Best Local Similarity: 22.66%      Mismatches: 216
Query Match:    9.05%      Indels:      196
DB:             7      Gaps:      35

US-09-750-986D-30 (1-589) x ADL23371 (1-2103)
Qy      9 P|r|o|p|h|e|P|r|o|g|y|A|r|g|l|e|u|g|l|u|A|r|g|l|e|u|g|l|u|l|e|t|r|p|a|l|a|y|h|r|a|r|g|P|r|o|g|l|u|g|n 28
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      241 C|T|G|A|A|A|T|G|G|G|A|C|T|T|A|C|A|T|G|A|C|A|A|T|T|T|G|T|A|C|A|T|G|C|T|T|G|A|A|C|A|T|A|T|G|T|G|A|A|A|C 300
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Qy      29 T|h|r|C|y|e|A|l|a|l|a|a|r|g|a|l|a|a|e|n-----G|y|g|l|u|t|r|p|r|a|r|g|l|e|s|e|r 44
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      301 A|A|G|A|T|C|T|T|G|T|A|C|A|G|A|G|T|G|C|G|A|T|G|A|A|C|A|T|T|G|A|G|A|G|A|T|C|A|T|G|A|T|C|A|G|A|T|C|A|G|A 360
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Qy      45 T|y|r|A|l|a|l|u|e|c|h|e|i|s|e|n|v|a|l|a|r|g|a|l|a|l|e|l|g|l|s|e|r|l|e|u|e|u|r|o|t|r|y|r|g|l|y|e|u 64
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      361 T|A|T|G|A|A|G|A|C|A|G|C|T|G|A|G|C|A|C|A|G|C|A|T|T|G|T|T|C|A|G|A|C|T|T|G|T|T|C|A|G|A|G|T 420
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Qy      65 S|e|r-----A|a|g|l|u|a|r|P|r|o|l|e|u|l|e|a|l|s|e|r|g|y 75
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      421 A|A|C|A|G|A|G|C|T|T|G|C|A|C|T|A|T|T|T|A|T|T|A|A|C|A|G|C|A|G|A|G|G|T|T----- 474
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Qy      76 A|e|n|a|p|l|e|u|g|l|l|e|u|g|l|u|e|a|l|a|p|h|e|g|y|a|l|e|u|e|r|y|a|l|g|l|y|l|e|p|r|o|t|r|y|C|y|e 95
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      475 -----G|T|G|A|T|C|A|T|C|T|T|G|C|A|G|C|A|T|A|T|C|A|T|T|G|T|C|T|G|T----- 513
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Qy      96 P|r|o|v|a|l|s|e|r|P|r|o|a|l|y|r|s|e|r|l|e|u|s|e|r|g|l|n|a|p|l|e|u|a|l|y|s-----L|e|u|A|r|G|H|s 113
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      514 -----C|T|T|A|T|A|T|A|T|A|C|A|C|T|T|G|T|C|A|G|C|G|T|T|A|A|G|T|T|G|G|A|A|T|C|A|T 564
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Qy      114 -----l|l|e|a|l|g|l|y|l|e|u|g|l|n|P|r|o|g|l|y|e|u|a|l|P|h|e|A|l 125
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      565 G|C|T|A|T|C|G|A|C|T|A|T|T|T|G|T|G|T|A|C|A|A|C|C|T|G|A|A|T|T|T|G|C|T|A|A|C|T|T|C|T|A 624
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Qy      126 A|l|a|s|p|a|l|a|l|p|r|o|h|e|g|l|n|a|r|g|a|l|l|e|g|l|u|r|h|l|e-----L|e|u|r|o|a|p|a|p|a|l 143
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      625 G|C|G|A|A|---A|T|C|C|A|T|C|A|T|T|G|C|T|C|A|T|T|G|T|G|T|G|G|A|G|G|G|C|G|T|G|A|G|C|A|T|T|G 681
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Qy      144 P|r|o|a|l|l|e|h|e|r|h|r|g|l|y|l|e|u|a|l|g|l|y|A|r|g|T|h|r|A|s|e|r|P|h|e|a|s|e|r|l|e|u 163
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      682 C|C|A|T|C|A|C|T|---C|C|T|G|A|G|A|-----A|C|T|G|A|G|T|C|A|A|T|T|G|T|A|T|C|A|T|C|A|A|A|A|G|C|T|A 732
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Qy      164 l|e|u|g|l|u|P|r|o|g|l|y|l|l|e|g|l|u|A|A|p|a|A|l|P|h|e|A|l|a|l|a|t|h|r|g|l|y|P|r|o|a|p|h|r 183
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      733 T|T|G|A|G|T|C|A|G|-----G|T|G|A|A|G|T|A|C|T|T|A|C|A|T|C|A|T|T|T|C|G|C|T|C|C|A|A|A|C|C|A|G|A|A|G|A|C 786
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Qy      184 l|l|e|a|l|y|a|p|h|e|u|h|e|r|h|r|s|e|r|g|l|y|s|e|r|h|r|L|y|e|u|P|r|o|l|y|a|l|a|v|a|l|P|r|h|r|h|r 203
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      787 A|T|T|G|C|A|C|A|T|A|T|G|T|A|C|A|C|A|G|T|G|A|C|A|C|A|G|A|C|C|A|A|A|G|G|T|G|T|G|T|G|A|C|T 846
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Qy      204 G|l|n|A|r|g|e|r|l|e|u|C|y|a|l|a|e|n-----G|l|n|G|l|e|r|l|e|u|e|u|l|n|h|r|P|h|e|P|r|o|v|a|l 220
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      847 C|A|T|G|A|A|A|C|T|T|G|C|G|A|A|T|G|C|G|T|G|T|C|A|G|T|G|G|A|A|G|C|A|A|A|T|T|C|T|C|T--- 903
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Qy      221 P|h|e|g|l|y|l|u|P|r|o|v|a|l|l|e|u|a|l|a|p|r|l|e|u|P|r|o|t|r|p|h|n|h|e|r|h|P|h|e|g|l|y 240
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

Novel acyl coenzyme A synthetases gene useful for altering a phenotype of

KW Genome; bacterium; Haemophilus influenzae; computer readable medium;
KW expression modulating fragment; regulation; gene expression; vector;
KW organism; open reading frame; ORF; de.
XX
OS Haemophilus influenzae.
XX
PN W09633276-A1.
XX
PD 24-OCT-1996.
XX
PF 22-APR-1996; 96WO-US005320.
XX
PR 21-APR-1995; 95US-00426787.
PR 07-JUN-1995; 95US-00476102.
XX 07-JUN-1995; 95US-00487429.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (UYJO) UNITV JOHNS HOPKINS.
PI Fleischmann RD, Adams MD, White O, Smith HO, Venter JC;
PI WPI, 1996-465782/48.
XX
DR Haemophilus influenzae Rd genome recorded on computer readable medium -
XX useful for identifying commercially important nucleic acid fragments by
XX PT homology searching.
XX
PS Claim 1; Page 77.2-77.1091; 1291pp; English.
XX
CC This sequence represents the complete genome sequence of the bacterium
CC Haemophilus influenzae strain Rd. The invention relates to a computer
CC readable medium (CRM) having recorded upon it the complete H. influenzae
CC nucleotide sequence (I), a representative fragment of (I) or a nucleotide
CC sequence at least 99% identical to (I). By providing the full-length
CC genomic sequence in a computer readable form, it is possible to identify
CC commercially important nucleic acid fragments and expression modulating
CC fragments (EMFs) of the Haemophilus genome. The EMFs can be used to
CC regulate the expression of a nucleic acid molecule. Vectors and altered
CC organisms comprising the predicted ORFs can be used to produce any of the
CC polypeptide fragments of the H. influenzae Rd genome
XX
SQ Sequence 1830121 BP; 567399A; 350615C; 347389G; 564036T; 0U; 6820cher;
XX
Alignment Scores:
Pred. No.: 3.17e-13 Length: 110000
Score: 283.00 Matches: 137
Percent Similarity: 38.44% Conservative: 114
Best Local Similarity: 20.98% Mismatches: 252
Query Match: 9.18% Indels: 150
DB: Gaps: 28
US-09-750-986d-30 (1-589) x AAT42063_00 (1-110000)
QY 15 LeuGluArgLeuGluHisTrpAlaValThrArgProGluGlnThrCysValAlaAlaArg 34
DB 646 GTTCATCGTATTCACCAACCAAGCCAAATCTCGTACAAATATGACCGCACTTCGTATAAA 705
QY 35 AlaAlaAsnGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 54
DB 706 ---GAACACGCGCTTGCGCAAGATATCTCTTGAAAACTTTCAAGCAACCAATCAAA 762
QY 55 AlaAlaGlnSerLeuLeuProGlyArgLysLysLysLysLysLysLysLysLysLys 74
DB 763 CTTTCTCGACGATTTGCTGCTCACAATATTTGACGTACAGATATAAATCGCATTTTGGCC 822
QY 75 GlyAsn-----AspLeuGluHisLeuGluLeuAlaPheGlyAla 87
DB 823 CATAAATATGGAAGCTTGACAAATCGTTGACATTCGACCTTACCAATTCGA----- 873
QY 88 MetTrpAlaGlyLysLeuProGlyCysProValSerProAlaTyrSerLeuLeuSerGlnAsp 107
DB 874 -----GCAATCACAGTGCCTATTTCACCAACCAATACAGCCAG 912

QY 108 LeuAlaLysLeuArgHisIleValGlyLeuLeuGlnProGlyLeuValPheAlaAlaAsp 127
DB 913 CAACGACAGATTT-----ATCCTTAATACGCGCATGTAATAAATCTCTTCTGGCGCAT 966
QY 128 AlaAlaProPheGlnAlaIleGluThrIleLeuProAspAspValProAlaIlePhe 147
DB 967 CAAAGCAATACGATCAACAATTCGAA-----ATTCTCATCTATTCCTCAAAATTA--- 1017
QY 148 ThrArgGlyLeuLeuAlaGlyArgThrValSerPheAspSerLeuLeuGlnPro 167
DB 1018 -----CAAAAATTGTACCAATGAAATCCACCATTCAA----- 1050
QY 168 GlyGlyIleGluAlaAsp-----AsnAlaPheAlaAlaThrGlyPro 181
DB 1051 -----TTACACAAAGATCTCTTCTTTCGACCTTGGAAAGTTTATTAAACAGTTCA 1104
QY 182 AspThr-----IleAla 185
DB 1105 AACGCCCAACAGATGAATTAACCAACGCTTAACCAAAACAAATTAATCGATTTATTT 1164
QY 186 LysPheLeuPheThrSerGlySerThrLysLeuProGlyAlaValProThrThrGlnArg 205
DB 1165 ACGATTATTTATACCTCAGCGACACGAGAGCTTAAAGGTTCATGATTAATACGCT 1224
QY 206 MetLeuCysAlaAsnGlnGlnMetLeuLeuGlnThrPheProValPheGlyGluGluPro 225
DB 1225 AATCTCGCT-----CACCAATTAGAAACACATCATCTTTCACCTTAATGACAGATCAG 1278
QY 226 ProValLeuValAspTrpLeuProTrpAsnHisThrPheGlyGlySerHisAsnIleGly 245
DB 1279 GATATTTCATCTTCTTTTTCATCTCTCAATTTTTCACGCGCATGG---GCGGCT 1335
QY 246 IleValLeuLysArgGlyGlyThrTrp---TyrLeuAspAspGlyLysProThrAlaGln 264
DB 1336 TATATTCTTATAGAGCGCAATATCTTGCTATTAAGAACAATAATCAAGTCGGGTCA 1395
QY 265 GlyPheAlaGluThrLeuArgAsnLeuSerGluIleSerProThrAlaTrpLeuThrVal 284
DB 1396 GCT-----TTACGGAATTCGCCCACTTAATATGTCGGCGCTA 1434
QY 285 ProLysGlyTrpGluLeuValGlyAla-----LeuGluArgAspSerThr 300
DB 1435 CCACGTTTTCAGAAAATAATTATGCTGCGTATGGAATAAGTCAAAAAGCCCAAAA 1494
QY 301 LeuArgGluArgPhePheAlaArgMetLysLeuPhePhePheAlaAlaGlyLeuSer 320
DB 1495 CTTGCGCA-----ATTATGTTTATTTGCGCAATTCCTCGTGGGA 1533
QY 321 GlnGlyIleTrpAsp---ArgLeuAspArgValAla----- 331
DB 1534 CAAAATATTTTATCTACGTGCGAATACAAAGCTATTCCTGTTTATGAAAGAAACAA 1593
QY 332 -----GlnGlnHisCysGlyGluArgIle 339
DB 1594 TTTGCTTTAGCAGATAAATTATGCTCTCAAAACTTCGTCAATTAATGCGGGGACGATA 1653
QY 340 ArgMetMet----- 342
DB 1654 AAAATGATGCTTCGCGAGAGCAAAATAGAAACCGGCTATTTGGCTATTTTCCATGCT 1713
QY 343 -----AlaGlyLeuGlyMetThrGluThrAlaPro-----SerCys 354
DB 1714 ATTGATATCAATCAATTAATGAGCTTATGATACGAAACAACCTGCAACCGTTTCTGC 1773
QY 355 -----ThrPheThrThrGlyProLeuSerMetAlaGlyTyrIleGlyLeuProAlaPro 372
DB 1774 TGGCAGATTTTCCAAATTAACCAATTC-----ATCGGACACATGATGCCA 1821
QY 373 GlyCysGlyValLysLeuValProValAspGlyLysLeuGluGlyArgPheHisGlyPro 392
DB 1822 AAAGCGGAAGTGAAA-----ATTGGGGAATAAATGAATTCCTTGTCGTGGCGGA 1872
QY 393 HisValMetSerGlyTyrTrpArgAlaProGluGlnAsnAlaGlnAlaPheAspGluGlu 412


```

Db      1315 TCGTTTGATGTTGCGTGCAGGATGATGACCGAGACT-----TCTTGTC 1368
Qy      356 PheThr-----GlyProLeuSerMetIaGlyTyrIleGlyLeuProAlaPro 372
      1369 ATAGTCTCTATGATGATGATGAC--AATTATCTGGCCATGTCGTTCCCTTAATCCA 1425
Qy      373 GlyCysGluValIleuValProValAsp----- 382
      1426 GCTTGCCAGGTAAACCTTGATGATGTTCCCGAATAATGATTAACATGACGATCAACCA 1485
Db      383 -----GlyIleuValGluGlyArgPheHisGlyProHisValMetSerGlyTyrTrp 399
      1486 TACCCACGTGTGTAATCTGTGAAGA-----GGACCAATCATCTTCMAAGGCTACTAC 1539
Qy      400 ArgAlaProGluIleuAlaGlnAlaPheAspGluGluGlyTyrTyrCysSerGlyAsp 419
      1540 AAGATGAAGAACCAACGACGAGAAATCTTGATGAGATGGCTGTAACACACAGAGAT 1599
Qy      420 AlaIleIleuValAspProAlaAspProGluIleuMetPheAspGly----- 437
      1600 ATC-----GGTTGTGTTACCTGTCGTCGG 1626
Db      438 -----ArgIleAlaGluAspPheIleuSerSerGlyValPheValSer 452
      1627 CTCAGATCATAGACAGAAAGAACATATTTAAGTTGGCGCAAGAGATATATAGCA 1686
Qy      453 ValGlyProLeuArg-----ThrArgAlaValLeuGlu 463
      1687 CCAAGAGAAATCGAAATGTTTATACAAATGATGATTCGTTTGCAGTGTTCATTAC 1746
Qy      464 GlyGlySerTyr-----ValIleuAspValValAlaAlaProAsp----- 477
      1747 GGGATATGCTTCAATTCCTCTAGTAGCTATAGTTTCAGTCGACCCCGAAGTTATGAA 1806
Qy      478 -----ArgGluCysLeuGlyLeuLeuVal---PhePro 487
      1807 GATTTGGCTCATCAGAACGATCAAGATGATGATGATGATGATGATGATGATGATGAT 1866
Db      488 ArgLeuLeuAsp-----CysArgAlaLeuSerGlyLeuGlyValIleu----- 502
      1867 AGAGTGCAGAAAGACTGTTCTTCTGAGATGATGATGATGATGATGATGATGATGATGAT 1926
Qy      503 -----SerAspAlaGluValIleuValAspGluProVal----- 513
      1927 GGGTTGAGTTTGCAGAAAGCTGTGATCTTGGTGCAGAACCATTTGAGAGATGGA 1986
Qy      514 -----ArgAlaTrpPheAlaAspTrp 520
      1987 CTCTCACACCAACATTCAGATTAAGAGACCTTCACCAAAAGCCTTCTTGACAGAGCA 2046
Qy      521 LeuIleuArgLeuAsnArgIleu 527
      2047 ATTAGCAAAATGATAGCGGAA 2067
Db
RESULT 13
ADN72468 standard; cDNA; 2103 BP.
AC      ADN72468;
DT      15-JUL-2004 (first entry)
DE      Thale cress cDNA upregulated in E2Fa/Dpa expressing plants Seqid 363.
KW      gene; ss; plant; transgenic; E2Fa/Dpa transcription factor;
KW      growth regulator; animal feed product; thale cress;
KW      cell wall biosynthesis; nitrogen metabolism; carbon metabolism.
OS      Arabidopsis thaliana.
PN      WO2004035798-A2.
XX

```

```

PD      29-APR-2004.
XX
XX      20-OCT-2003; 2003MO-EP011658.
XX
PR      18-OCT-2002; 2002EP-00079408.
XX
PA      (CROP-) CROPEDESIGN NV.
XX
PI      Inze D, De Veylder L, Vlieghe K;
XX
DR      WPI: 2004-348456/32.
XX
XX      P-PSDB; ADN72469.
XX
PT      Altering plant characteristics, useful for producing plants for enzyme or
PT      pharmaceutical production comprises modifying in a plant, expression of
PT      one or more nucleic acids and/or modifying level or activity of one or
PT      more proteins.
XX
XX      Claim 1; SEQ ID NO 363; 134pp; English.
XX
CC      This invention relates to a novel method for altering one or more plant
CC      characteristics. Specifically, it refers to identifying genes that are up
CC      - or down-regulated in transgenic plants overexpressing the heterodimeric
CC      E2Fa/Dpa transcription factor of Arabidopsis and using these sequences to
CC      alter plant characteristics accordingly. The present invention describes
CC      generating transgenic plants for the production of growth regulators,
CC      enzymes, therapeutics, pharmaceuticals and animal feed products, where
CC      the altered plant characteristics are selected from increased yield or
CC      biomass, enhanced survival capacity, stress tolerance, plant architecture
CC      or physiology, altered endoreduplication, biochemistry, signal
CC      transduction, storage lipid mobilization and/or altered photosynthesis,
CC      each relative to the corresponding wild type plants. Accordingly, these
CC      sequences can also be useful as positive or negative selectable markers
CC      during transformation of cells or tissues. The identified genes play a
CC      role in a variety of biological processes such as DNA replication, cell
CC      wall biosynthesis, nitrogen and/or carbon metabolism or they function as
CC      transcription factors. This polynucleotide sequence is thale cress cDNA
CC      upregulated 1.3 fold or more in plants overexpressing the E2Fa/Dpa
CC      transcription factor, given in an exemplification of the invention.
XX
SQ      Sequence 2103 BP; 589 A; 418 C; 496 G; 600 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 1,49e-15 Length: 2103
Score: 283.00 Matches: 150
Percent Similarity: 37.03% Conservative: 97
Best Local Similarity: 22.49% Mismatches: 214
Query Match: 9.18% Indels: 206
DB: Gaps: 35
US-09-750-986D-30 (1-589) x ADN72468 (1-2103)
Qy      9 ProPheProGlyArgIleuGluArgLeuGluHisTrpAlaIleuThrArgProGluGln 28
      241 CCGAATTCGGACCTTACATGACAAATTTGTACATGCTGTGTAACATGCTGAAC 300
Qy      29 ThrCysValAlaIleuArgAlaIleuAsn-----GlyGluTrpArgArgIleSer 44
      301 AAGTATCTTGATACAGAGTTCGTCGATGACCAATTTGAGAGATTTATGATGATACA 360
Qy      45 TyrAlaGluMetPheHisAsnValArgAlaIleuAlaGlnSerIleuLeuProTyrGlyLeu 64
      361 TATGAGAAACAGCGTGTGAGCAAGCCATTTGTTGACAGACTCTTGTTCATGAGGTT 420
Qy      65 Ser-----AlaGluArgProLeuLeuIleuValSerGly 75
      421 AACCAAGAGATTGCGTTGACTCTATTATTATTAACGACAGAGTGTTGTT----- 474
Db      76 AsnAspLeuGluHisLeuGluIleuAlaPheGlyAlaMetTyrAlaGlyIleProTyrCys 95
      475 -----GTGATCATGCTTGTGACGCAATATTCATTTGTCTCTGTT----- 513
Qy      96 ProValSerProAlaTyrSerLeuLeuSerGlnAspLeuAlaIleu-----LeuArgHis 113

```


WP AC67367_54 540001 551000
WP AC67367_55 550001 561000
WP AC67367_56 560001 564894

Alignment Scores:
Pred. No.: 7.33e-14 Length: 110000
Score: 290.00 Matches: 146
Percent Similarity: 40.13% Conservation: 110
Best Local Similarity: 22.88% Mismatches: 264
Query Match: 9.41% Indels: 118
Gaps: 26

US-09-750-986d-30 (1-589) x AC67367_48 (1-110000)

QY 11 ProGlyArgIleuGluArgLeuGluHieTrpAlaIysThrArgProGluGlnThrCys 30
Db 59870 CCTTATCACTTGATTAACCGGTTGCAACACAAATMAAACACCCAGAGAAA----- 59923

QY 31 ValAlaAlaArgAlaAlaAsnGlyGluTrpArgArgIleSerTrpAlaGluMet---Phe 49
Db 59924 ATTGCTTTTGGC-----CAATGGCTCCGGCTGAAACACGAACTGAACCTG 59971

QY 50 HieAsnVal-----ArgAlaIleAlaGlnSerLeuLeuProTrpGlyLeuSer 65
Db 59972 CATGATATTGGCAAAAAGACCGACGCTATTAGCCGGCGTGTGGAGATTGGTGTGAT 60031

QY 66 AlaGluArgProLeuLeuIleValSerGlyAsn-----AspLeu 78
Db 60032 GTGCACAAAAAATTGGCTCTTTGGCACATBACAGCATGGCGTGTCACTGGCAGATATC 60091

QY 79 GluHieLeuGlnLeuAlaPheAlaIlePheGlyAlaMetTrpAlaGlyIleProTrpCysProValSer 98
Db 60092 GCTGATTACAACTTGG-----GGGTCACATGTC 60121

QY 99 ProAlaTrpSerLeuLeuSerGlnAspLeuAlaIysLeuArgHieIleValGlyLeuLeu 118
Db 60122 CCTCTGATCCACACCACTAGTCAGGAAACAGCGTGGTTTC-----ATACTTAATGATGCT 60175

QY 119 GlnProGlyLeuValPheAlaAlaAspAlaAlaProPheGlnArgAlaIleGlu---Thr 137
Db 60176 GGTGTGGCGATTTTGTGTGGTGGGACGGCAGATGATGTTGCTATTACCTTGACA 60235

QY 138 IleLeuProAspValProAlaIlePheThrArgGlyLeuLeuAlaGlyArgArgThr 157
Db 60236 GAATGTTGCCCTCAGTTAAGTCATATTGTCTTAAGATGATGACGATTTGGCGGAT 60295

QY 158 ValSerPheAspSerLeuLeu-----GluGlnProGlyIle 170
Db 60296 TGTTCATTACCTCAGCATTTATCTGATTCATGTCGATGATGAATTAACCTGTCACTCC 60355

QY 171 GluAlaAspAsnAlaPheAlaIleThrGlyProAspThrIleAlaIysPheLeuPheThr 190
Db 60356 GAGTTGGAACCTGATTATGCTAACACCAATCTGATATCTTTTACACTTATTATTAACG 60415

QY 191 SerGlySerThrIysLeuProIysAlaValProThrThrGlnArg-----Met 206
Db 60416 TCTGGTACCAACGGGTGAGCCTAAAGCGTCATGCTTATTAACCCGATTTAGCCTATCAG 60475

QY 207 LeuCysAlaAsnGlnMetLeuLeuGlnThrPheProValPheGlyGluGluProPro 226
Db 60476 TTTGATTACATGATCATATCTCTGGCATTTAACT-----GATCAGAC 60517

QY 227 ValLeuValAspTrpLeuProTrpAsnHieThrPheGlyGlySerHieAsnIleGlyIle 246
Db 60518 ATTTCCTTGGTTTCTGCGGTTATCCCATGCTTTTACGCGTGATGAGATGCTTATGTC 60577

QY 247 ValLeuTrpAsnGlyThrTrpTrpLeuAspAspGlyIysProThrAlaGlnGlyPhe 266
Db 60578 ATGCATGTAGGGGACAAATGTTTATCTGACGACACTAATTTGTCTGCTCAAGCATAG 60637

QY 267 AlaGluThrLeuArgAsnLeuSerGluIleSerProThrAlaTrpLeuThrValProIys 286
Db 60638 GCG-----GATGTCGCCCGACAGTGAATGATGTCGATGCCGAGA 60676

QY 287 GlyTrpGluLeuValGlyAlaLeu---GluArgAspSerThrLeuArgGluArgPhe 305
Db 60677 TTTTACGAAAAAATTATAGTCGGCTTTTGAAAAGCTTTCT-----AAGCTCCT 60727

QY 306 PheAlaArgMetIysLeuPhePheAla---AlaAlaGlyLeuSerGlnGlyIle--- 323
Db 60728 TGGCATCGAAGAAAACCTTTTAATTGGCGCATTTAAATGTGGTGAGCGCTGATTTTACGT 60787

QY 324 -----TrpAsp---ArgLeuAspArgValAlaGluGln 333
Db 60788 AAGTGCAGATAAAAAACCTGCCCTTGGTGCACACCAATGACATGCGCTGATPAA 60947

QY 334 -----HieCysGlyGluArgIleArgMetAla----- 343
Db 60848 TTGGCTTGACCAAAATTACGGAATGTGTGGGGGACAGATGCGTTTTTACCTGCTCT 60907

QY 343 ----- 343
Db 60908 GGTCACGCTGACGAAACGATTATCTGTTCTTCTGCTGTGGGATPAAATATCAAA 60967

QY 344 ---GlyLeuGlyMetThrGluThrAlaPro-----SerCysThrPheThrThrGlyPro 360
Db 60968 TATGCTATGTCATGACAGAGACTTGTGCCACGGTTTCTGTGGAAAGCAAGT--- 61024

QY 361 LeuSerMetAlaGlyTrpIleGlyLeuProAlaProGlyCysGluValIysLeuValPro 380
Db 61025 ---TATGCTACTGGTTTATTTGTTAAACATCCGCTGGCATTTACGTCCTGATCGGAA 61081

QY 381 ValAspGlyIysLeuGluGlyArgPheHieGlyProHieValMetSerGlyTrpTrpArg 400
Db 61082 GAAGAT-----GAGATACAAAGTCCGTGCCCAATTTGATGAAGGCTATTTCAT 61132

QY 401 AlaProGluGlnAsnAlaGlnAlaIlePheAspGluGluGlyTrpTrpCysSerGlyAspAla 420
Db 61133 AATCAGCAAAACCAACAGCAATTTACTCAAGACGGTGGTGCCTACAGGAGTGC 61192

QY 421 IleIysLeuAlaAspProAlaAspProGlnIysGlyLeuMetPheAspGlyArgIleAla 440
Db 61193 GGTAACTG-----GATGAAATGCTAATTTGCTACTTATGAAACGCTGAAA 61240

QY 441 GluAspPheIysLeuSerSerGlyValPheValSerValGlyProLeuArgThrArgAla 460
Db 61241 GATTTGATGAACATCATGAAATGCAATATATATGCG-----CCTCAATGATGAAGGC 61294

QY 461 ValLeuGluGlyGlySerTrpValIleAspValValAlaAlaProAspArgGluCys 480
Db 61295 ATTTTGGACAAAGATCGTTTATTGAAACATGTTGCTATTATTTGCTGATGCTGTAATTT 61354

QY 481 LeuGlyLeuLeuValPhePro-----ArgLeuLeuAspCysArgAlaLeuSerGlyLeu 498
Db 61355 GTATCGGCTTTTATTCGCTTGTCTCATGAAGCATTTGGAAGAGTATGCAAGTCAATTAAT 61414

QY 499 GlyIysGluAlaSerAspAlaGluValIleuAlaSerGluProValArgAlaTrpPheAla 518
Db 61415 CTCAAATATCATGACCGGCTGTAATTAATCTGTCATGATGATGATCAATATGTTGAG 61474

QY 519 AspTrpLeuIysArgLeuAsnArgGluAlaThrGlyAsnAlaSerArgIleMetTrpVal 538
Db 61475 CAACGCTTGAAATAATGACGAAAC---TTTGTCTCACTGCATGACGTAAGCGTTTC 61531

QY 539 GlyLeuLeuAspThrProProSerIleAspIysGlyGluValThrAspIysGlySerIle 558
Db 61532 ATATTACTGCCAAAACCATTTTCAATGGAACGGGTGAATTAATCTCAACACTGAAATTTG 61591

QY 559 AsnGlnArgAlaValLeuGlnTrpArgSerAlaIysValAspAlaLeuTrpArg 576
Db 61592 AGAGGAAATTTATTATGACGGTTTCTATGAAGAAATTGAATCTATGATCAG 61645

RESULT 12
ADL72353
ID ADL72353 standard; cDNA; 2103 BP.
XX

```
Db 568 TCTGATACCCGCGGTGAGCCTTAAGCGGTATGCTTATTCACCGCAATTTAGCCTATCAG 627
Qy 207 LeuCyAlaAsnGlnMetLeuLeuGlnThrPheProValPheGlyGluProPro 226
Db 628 TTGTATTCATCATGATCATCTCTGCGCATTAATCT-----GATCAGGAC 669
Qy 227 ValLeuValAspThrLeuProThrPheAsnIleThrPheGlyGlySerHisAsnIleGlyIle 246
Db 670 ATTCTTTGAGTTTCTGCGGTTATCCCATGCTTTTGAAGCGGTGAGGATGATGATCTATGTC 729
Qy 247 ValLeuTyraAsnGlyGlyThrTyrTyLeuAspAspGlyLysProThrAlaGlnGlyPhe 266
Db 730 ATGCATGTAGGGGACAAATAATGTTATCTGACGTGACACTAATTTGTCTGCAAGCTATG 789
Qy 267 AlGluThrLeuArgAsnLeuSerGluIleSerProThrAlaTyrLeuThrValProLys 286
Db 790 GCG-----GATGCTCGCCGACAGTGTGATGTCGACGTGCCGAGA 828
Qy 287 GlyThrGluGluLeuValGlyAlaLeu---GluArgAspSerThrLeuArgGluArgPhe 305
Db 829 TTTTACGAAAAATTATTAAGTTCGCGTTTGAAGGTTTCT-----AAGCTCCT 879
Qy 306 PheAlaArgMetLysLeuPhePhePheAla---AlaAlaGlyLeuSerGlnGlyIle--- 323
Db 880 TGGCATGGAAGAAACTTTTAAATGCGCATTAATGCGTGAAGCTCAGTTTAACT 939
Qy 324 -----TyrAsp---ArgLeuAspArgValAlaGluGln 333
Db 940 AAGTTCAGAGTAAAAAAGCTCCCTTGGTCAACAGCAATGCATCAGCTGATGATAA 999
Qy 334 -----HisCysGlyGluArgGlyLeuArgMetMetAla----- 343
Db 1000 TTGGTGTAGCAAAATTACGAAATGTGTGGGCGGACAGTGTGCTTTTAACTGCTGCT 1059
Qy 343 ----- 343
Db 1060 GGTGACGCTGACGAAAGATTAATCTGTTCTTTCGCTGTTGGGATTAATGCAA 1119
Qy 344 ---GlyLeuGlyMetThrGluThrAlaPro---SerCysThrPheThrGlyPro 360
Db 1120 TATGCTATGAGTGAAGAGACTTGTGCAAGGTTTCTGTTGGGAAAGCAAGGT--- 1176
Qy 361 LeuSerMetAlaGlyTyrlleGlyLeuProAlaProGlyCysGluValLysLeuValPro 380
Db 1177 ---TATGCTACTGTTCTATGTTGAACCAATCGCTGCAATGACGTCCGATCGGTGAA 1233
Qy 381 ValAspGlyLysLeuGluGlyArgPheHisGlyProHisValMetSerGlyTyrrPrArg 400
Db 1234 GAAGAT-----GAGATACCAAGTCCGTGCGCCCATTTGTGATGAAGGCTATTTCAAT 1284
Qy 401 AlaProGluGlnAsnAlaGlnAlaPheAspGluGlyTyrrTyCysSerGlyAspAla 420
Db 1285 AATCAGCAAGAAACCAACCAAGGCAATTAATCAAGACGTTGTTGCTGACAGGTATGCC 1444
Qy 421 IleLeuLeuAlaAspProAlaAspProGlnLysGlyLeuMetPheAspGlyArgIleAla 440
Db 1345 GGTAAAGTTG---GATGAAATGTGAATTTGTCATTACTGAACGCTGAAA 1392
Qy 441 GluAspPheLysLeuSerSerGlyValPheValSerValGlyProLeuArgThrArgAla 460
Db 1393 GATTGATGAAGAACATGCAATGCAATATATTCG-----CCTCAATGATGGAAGGC 1446
Qy 461 ValLeuGluGlyGlySerTyrrValLeuAspValValAlaAlaProAspArgGluCys 480
Db 1447 ATTTTGGCAAGATCGTTTATTAAGACATGTTGCTTATTTGCTGATGCTCGTAAATTT 1506
Qy 481 LeuGlyLeuLeuValPhePro-----ArgLeuLeuAspCysArgAlaLeuSerGlyLeu 498
Db 1507 GTATCGGCTTATCGCTTGTGATGCAAGACATGGAAGAGTGAAGGTCATTAAT 1566
Qy 499 GlyLysGluAlaSerAspAlaGluValLeuAlaSerGluProValArgAlaTrpPheAla 518
```

```
Db 1567 CTCAAATATCATGACCGGCTGAAATTACTGCGTCATAGTCAGATATCATATGTTGAG 1626
Qy 519 AspThrLeuLysArgLeuAsnArgGluAlaThrGlyAsnAlaSerArgIleMetThrVal 538
Db 1627 CAACGCTTGAAGAAATATGAGAAAC---TTTGTCACTTCATCATGATGATGAAGCGTTTC 1683
Qy 539 GlyLeuLeuAspThrProPheSerIleAspLysGlyValValThrAspLysGlySerIle 558
Db 1684 ATATTACTGCAAAACATTTTCAATGTAAGGAGGATTAATCAACACAGTGAATTC 1743
Qy 559 AsnGlnArgAlaValLeuGlnThrPargSerAlaLysValAspAlaLeuTyraArg 576
Db 1744 AGAGAGAAATATTAATGACGCTTACTATGAAGAAATTAATGATGATATCAG 1797

RESULT 11
ACF67367_48
Continuation (49 of 57) of ACF67367 from base 4800001 (Photobhabdus luminescens nucleotic
WP Sequence split into 57 fragments Locus ACF67367 Accession ACF67367
WP Fragment Name Begin End
WP ACF67367_00 1 110000
WP ACF67367_01 100001 210000
WP ACF67367_02 200001 310000
WP ACF67367_03 300001 410000
WP ACF67367_04 400001 510000
WP ACF67367_05 500001 610000
WP ACF67367_06 600001 710000
WP ACF67367_07 700001 810000
WP ACF67367_08 800001 910000
WP ACF67367_09 900001 1010000
WP ACF67367_10 1000001 1110000
WP ACF67367_11 1100001 1210000
WP ACF67367_12 1200001 1310000
WP ACF67367_13 1300001 1410000
WP ACF67367_14 1400001 1510000
WP ACF67367_15 1500001 1610000
WP ACF67367_16 1600001 1710000
WP ACF67367_17 1700001 1810000
WP ACF67367_18 1800001 1910000
WP ACF67367_19 1900001 2010000
WP ACF67367_20 200001 2110000
WP ACF67367_21 2100001 2210000
WP ACF67367_22 2200001 2310000
WP ACF67367_23 2300001 2410000
WP ACF67367_24 2400001 2510000
WP ACF67367_25 2500001 2610000
WP ACF67367_26 2600001 2710000
WP ACF67367_27 2700001 2810000
WP ACF67367_28 2800001 2910000
WP ACF67367_29 2900001 3010000
WP ACF67367_30 3000001 3110000
WP ACF67367_31 3100001 3210000
WP ACF67367_32 3200001 3310000
WP ACF67367_33 3300001 3410000
WP ACF67367_34 3400001 3510000
WP ACF67367_35 3500001 3610000
WP ACF67367_36 3600001 3710000
WP ACF67367_37 3700001 3810000
WP ACF67367_38 3800001 3910000
WP ACF67367_39 3900001 4010000
WP ACF67367_40 4000001 4110000
WP ACF67367_41 4100001 4210000
WP ACF67367_42 4200001 4310000
WP ACF67367_43 4300001 4410000
WP ACF67367_44 4400001 4510000
WP ACF67367_45 4500001 4610000
WP ACF67367_46 4600001 4710000
WP ACF67367_47 4700001 4810000
WP ACF67367_48 4800001 4910000
WP ACF67367_49 4900001 5010000
WP ACF67367_50 5000001 5110000
WP ACF67367_51 5100001 5210000
WP ACF67367_52 5200001 5310000
WP ACF67367_53 5300001 5410000
```

```
QY 451 ValSerValGlyProLeuArgThrArgAlaValLeuGluGlySerTyrValLeuAsp 470
Db 10512 ATTCGCTCAGCGCATCGAG-----AATTCGCTCATGAAGAC 10550
QY 471 -----ValValAlaAlaProAspArgGluCysLeuGlyLeuVal 485
Db 10551 CCCCTCTCGAACACGGCGGTGCTGGCGGATACCGTCCGTGCTGACTCTGCTGTT 10610
QY 486 PheProArgLeuLeuSerCysArgAlaLeu-----SerGlyLeu 498
Db 10611 AAGCCGCTGCTCCTCAGTGAAGAGATTGGCCGAGAGTTGCATACCTCATATACG 10670
QY 499 GlyValGluAlaSerAspAlaGluValLeuAlaSerGluProValArgAlaTyrPheAla 518
Db 10671 GGCCTCGGAATGCTGCTCCGAGAGACTAGCGGAGAA---ATTGCTCGCGCGGTGCC 10727
QY 519 AspTyrLeuValArgLeuAsnArgGluAlaThrGlyValAsnAlaSerArgGlyMetTyrVal 538
Db 10728 GAGATCACCGAGAAATTCCTCATCAGACAGATTGCTGATCTGACGAGTGTGTGGAC 10787
QY 539 GlyLeuLeuAspThrProProSerIleAspLysGlyGluValThrAspLysGlySerIle 558
Db 10788 GAGTTC-----ACCACAGACACGAGACTGCTCACTCCAACCTTAAGATA 10832
QY 559 AsnGlnArgAlaValLeuGlnTyrPargSerAlaLysValAspAlaLeuTyrArgGly-- 577
Db 10833 CGCCGACGTAGGTGAGAAAGATTCCACCGAGATCGTTGAGAGATGTACGCGGCTT 10892
QY 578 ----GluAspGlnSerMetLeuArgAsp 585
Db 10893 GCGGCGCGACGAAAGGCTGGTAAAGAG 10920

RESULT 10
ACF71666
ID ACF71666 standard; DNA, 1803 BP.
AC ACF71666;
XX
XX
XX 20-NOV-2003 (first entry)
XX
XX Photorhabdus luminescens nucleotide sequence #10133.
XX
XX Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
XX detection; food; gene expression; plant; animal; microorganism; toxin;
XX antibiotic; biopesticide; virulence factor; disease model; plague;
XX whooping cough; gene; ds.
XX
XX Photorhabdus luminescens.
XX
XX MO200294867-A2.
XX
XX 28-NOV-2002.
XX
XX 07-FEB-2002; 2002MO-IB003040.
XX
XX 07-FEB-2001; 2001FR-00001659.
XX
XX (INSP ) INST PASTEUR.
XX (CNRS ) CNRS CENT NAT RECH SCI.
XX
XX Duchaud E, Taourit S, Glaeser P, Frangeul L, Kunst P, Danchin A;
XX Buchrieser C;
XX
XX WPI; 2003-148459/14.
XX
XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
XX useful e.g. as therapeutic antimicrobials and agricultural pesticides.
XX
XX Claim 2; SEQ ID NO 10133; 1205pp; French.
XX
XX The invention relates to the isolation of genes and their encoded
XX proteins from Photorhabdus luminescens. The isolated sequences are
XX
```

```
CC sources of probes and primers for detecting the genome of P. luminescens
CC and related species; to study polymorphisms; for gene analysis and for
CC detection/amplification of the genes. Antibodies (Ab) raised against the
CC polypeptides encoded by the genes are used for detection/identification
CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
CC carry a gene-containing vector are used to select compounds that
CC modulate, regulate, induce or inhibit expression of the genes in plants,
CC animals or microorganisms other than P. luminescens and are able to alter
CC response or sensitivity to toxins and antibiotics produced by P.
CC luminescens. Cells transformed to express the genes are useful for
CC recombinant production of the proteins, particularly toxins and
CC antibacterials useful as insecticides, bactericides and fungicides. The
CC genes, proteins, vectors containing the genes and Ab are also useful
CC therapeutically (to treat microbial infection by bacteria or fungi that
CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as
CC biopesticides. Other uses of the genes and the proteins are as virulence
CC factors and for identifying targets of human diseases for which P.
CC luminescens is a model (particularly plague and whooping cough). This
CC sequence represents one of the isolated P. luminescens genes
XX
XX SQ Sequence 1803 BP; 502 A; 324 C; 423 G; 554 T; 0 U; 0 Other;
XX
XX Alignment Scores:
XX Pred. No.: 2.79e-16 Length: 1803
XX Score: 290.00 Matches: 146
XX Percent Similarity: 40.13% Conservative: 110
XX Best Local Similarity: 22.88% Mismatches: 264
XX Query Match: 9.41% Indels: 118
XX DB: 10 Gaps: 26
XX
XX US-09-750-986D-30 (1-589) x ACF71666 (1-1803)
QY 11 ProGlyArgIleuGlnArgLeuGluIleTyrAlaLeuThrArgProGluGlnThrCys 30
Db 22 CTTATCAGCTTGTATTAACCGGTTCGCAACAAATAAACAACCCAGAGAA----- 75
QY 31 ValAlaAlaArgAlaAlaAsnGlyGluTyrPargArgIleSerTyrAlaGluMet---Phe 49
Db 76 ATTGCTTTTTCG-----CAATGCTTCGCGCTGACAACTCGAACTTAAGTGG 123
QY 50 HisAsnVal-----ArgAlaIleAlaGlnSerLeuLeuProTyrGlyLeuSer 65
Db 124 CATGATTTTGGCAAAAGACCGACGAGCTATTAAGCGGCGTGTGTGAGATTGGTGTGAT 183
QY 66 AlAGluArgProLeuLeuIleValSerGlyAsn-----AspLeu 78
Db 184 GTCCAGAAAGAAATTTGGCTCTTTCGACATTAACAGATGCGGTGATTTGGCAGATATC 243
QY 79 GluHisLeuGlnLeuAlaPheGlyAlaMetTyrAlaGlyIleProTyrCysProValSer 98
Db 244 GCTGTATTACAACTTCG-----GCGGTCACTGTC 273
QY 99 ProAlaTyrSerLeuLeuSerGlnAspLeuAlaLysLeuArgHisIleValGlyLeuLeu 118
Db 274 CCTGTGTATGCAACAGTACTCAGAACAGCGGTTC-----ATACTTAATGATGCT 327
QY 119 GlnProGlyLeuValPheAlaAlaAspAlaAlaProPheGlnArgAlaIleGlu---Thr 137
Db 328 CGTGTGCGATTTTGTGTTGTGGGCGAGCCGACGATGATGTGCTATTAGCTTACACA 387
QY 138 IleLeuProAspAspValProAlaIlePheThrArgGlyGluLeuAlaGlyArgThr 157
Db 388 GAATGTTCCTCCAGTTAAGTCAATTAATGTCTTAGATGAATCAGTCGATTCGGGAT 447
QY 158 ValSerPheAspSerLeuLeu-----GluGlnProGlyGlyIle 170
Db 448 TGTTCATTAGCTCAGCATTTATCTGATTTTCATGTCGATGATTAATTAACCTTCAGTCC 507
QY 171 GluAlaAspAsnAlaPheAlaAlaThrGlyProAspThrIleAlaLysPheLeuPheThr 190
Db 508 GAGTTGGAACCTGTTATGTCTAACACAAATCTGGATGATCTTTTACACTTATTATACG 567
QY 191 SerGlySerThrLysLeuProLysAlaValProThrThrGlnArg-----Met 206
```

PT or for stimulating an immune response specific for a P. acnes protein.
 XX Claim 1, SEQ ID NO 17, 1481bp, English.

CC The invention relates to an isolated polynucleotide (ACF64435-ACF64733)
 CC encoding a Propionibacterium acnes protein. The invention also relates to
 CC polypeptides encoded by the polynucleotide (ABM35624-ABM64536) and to
 CC immunogenic fragments of P. acnes polypeptides. The invention
 CC additionally encompasses expression vectors and host cells comprising a
 CC polynucleotide of the invention; antibodies against polypeptides of the
 CC invention; fusion proteins comprising a polypeptide of the invention; a
 CC method for stimulating an immune response specific for a P. acnes
 CC polypeptide and an isolated T cell population comprising T cells prepared
 CC via this method; a vaccine composition (comprising P. acnes polypeptides,
 CC polynucleotides, antibodies, fusion proteins, T cell populations, or
 CC antigen-presenting cells that express the polypeptide); a method and kit
 CC for detecting or determining the presence or absence of P. acnes in a
 CC patient; and a method for inhibiting the development of P. acnes in a
 CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion
 CC proteins, T cell populations or antigen-presenting cells that express the
 CC polypeptides are useful for diagnosing, preventing or treating acne
 CC vulgaris, or for stimulating an immune response specific for a P. acnes
 CC protein. The polynucleotides can also be used as probes or primers for
 CC nucleic acid hybridization. The vaccine composition is useful for the
 CC stimulation of an immune response against P. acnes, or for treating acne,
 CC and the kit is useful for performing a diagnostic assay. The present
 CC sequence represents a P. acnes DNA contig which is specifically claimed
 CC in the invention. Note: The sequence data for this patent did not form
 CC part of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 37286 BP, 7034 A, 10065 C, 11922 G, 8264 T, 0 U, 1 Other:

Alignment Scores:

Pred. No.:	1,52e-15	Length:	37286
Score:	301.50	Matches:	154
Percent Similarity:	39.10%	Conservative:	108
Best Local Similarity:	22.99%	Mismatches:	264
Query Match:	9.78%	Indels:	144
DB:	8	Gaps:	29

US-09-750-986D-30 (1-589) x ACP64451 (1-37286)

QY 9 PropheProgluArg-----lleuGlulArgleuGluHlstrPalaVrThr----- 24
 DB 9066 CCGTACCGGGGAGAACTTCATTGAGAGCCACTCCGCCACATGTTCCGGCCACGGTT 9125
 QY 25 -----ArgProgluGlnThrCyValAlaAlaArgAla---AlaAsnGly 38
 DB 9126 GCCAACCCAGGGTTCCTCC-----GCGACTCGGGTTCCTCCAGGGCGGT 9170
 QY 39 GluTrpArgArgIleSerTyraIaGluMeIpheHisAnValArgAlaIleAlaGlnSer 58
 DB 9171 CAGTGTGATTCGCGACCTATGCGGAGACTGGCGGTGCGTAGCGGGCTCGCCGGGCC 9230
 QY 59 LeuLeuProTyrluLeuSerAlaGluArgProleu-----LeuIle 72
 DB 9231 TTGTGACTCCCGGGCTGCTCACCGAGATGCGCTGAGCGCGGGATGCTATTCTCTG 9290
 QY 73 ValSerGlyAnAspLeuGluHlstrPalaVrThrCyValAlaAlaArgAlaIleAlaGln 92
 DB 9291 TTGGCCGCTAAATTCGCCGAGGTGATCGAGGCTGATTCGCCGCGCATGAGATGGGCTG 9350
 QY 93 ProTyCyAspProValSerProAlaTySerleuLeuSerGlnAspLeuAlaValLeuArg 112
 DB 9351 -----ATCCCGGGT---CCGATTACCCGACGTCGACCCCTGAT-----CAGATTGTT 9395
 QY 113 HisIleValAlaGluLeuGlnProGlyLeuValPheAlaAlaAspAlaAlaProPheGln 132
 DB 9396 CACATCTCTCACTGATGCTGGGGTGGCTGCTATCATCAAGGCTGCTCCCAAGAGACTCGAC 9455
 QY 133 ArgAlaIleGluThrIleLeuProAspAspValProAlaIlePheThr----- 148
 DB 148

DB 9456 CGTATTCGAGGCG-----GCGCATCAGATGCCGGCCTTGAGACGGATCTGTATA 9509
 QY 149 -----ArgIlyleuAlaGlyArgArgThrValSerPheAspSerleu 153
 DB 9510 AATCCGAGATCAGGTTGGTGTATGATGACGGTTTGACGGGTCTGCTCCGTGAACAG--- 9566
 QY 164 LeuGluGlnProGlyIleGlyIle-----AlaAspAnAlaPheAla 178
 DB 9567 GTCCGTACGGCTGGGTTTCGAGGAGATACAGACGGTGTGAGAGCGCATGGGGCAG 9626
 QY 179 ThrGlyProAspThrIleAlaValPheLeuPheThrSerGlySerThrTyValLeuProGly 198
 DB 9627 TCAATGCCGAGCATGTTGGCCGTTGATTTACCTCCGGAACCAAGCGGTGAGCCCAAG 9686
 QY 199 AlaValProThrThrGlnArgMetLeuCyAlaAsnGlnMetLeuLeuGlnThrPhe 218
 DB 9687 GGGGTATGATCAAGCATCGGGCGGGCGGTGGCTGAG-----CTGCGAGGACATT 9734
 QY 219 ProValPheGlyIleGluGlnProVal-----LeuValAspTrpLeuProTrpAnHis 236
 DB 9735 GACGCTTCTTCGACGTACACCCAGCTGATCACTCGCTAGCTTCTGCGCTCTGCGAC 9794
 QY 237 ThrPheGlyIleSerHisAnIleGlyIleValLeuTyraAnGlyIleTyrluLeu 256
 DB 9795 GCCCTGAGTGGGAGATGGATGCGGATGATGCAATGTTGCTTCATACCTTCGTG 9854
 QY 257 AspAspGlyLeuProThrAlaGlnGlyPheAlaGluThrLeuArgAnLeuSerGlyIle 276
 DB 9855 CCACACCCGAAACGATTCG-----GCAATGTTGGCTGAGGTC 9893
 QY 277 SerProThrAlaSerThrValProGlyIleTyrluGluLeuValGlyAlaLeuGlu 296
 DB 9894 CGCCCAACCTTTTGTCTCAGTCCGAAACTGATGAAACATGATG----- 9941
 QY 297 ArgAspSerThrLeuArgGluArgPhe-----AlaArgMetLeuPhePhe 313
 DB 9942 -----AGCGTGGCCGTGAGAAAGTTTGTGATTCGCGCAAGCTCAAAATCTTTGAG 9995
 QY 314 PheAla----- 315
 DB 9996 TGGTCATACGATGGCCGCGAGTGTGCAAGCGAGACGAGAGGAGCGGCCCGAC 10055
 QY 316 ---AlaAlaGlyLeuSerGlnGlyIleTrpAspArgLeu----- 327
 DB 10056 GTATCTTGGAGACGGCTATGAGGTTGCCGATGCTGTCTCTCAAGGCGCATCCGTGAC 10115
 QY 328 -----AspArgVal 330
 DB 10116 GCCATGGGTGACCCAAAGACTGTGCTGGCGCGGAGCTCCATGCGCAAGAGTTC 10175
 QY 331 AlaGluGlnHisCyAspGlyIleArgIleArgMetAlaGlyLeuGlyMetThrGluThr 350
 DB 10176 GAAGAGTTCCTCCGCCCTCCGGCTGTGGTGTGSCAAGGAAATGCGGCTGACAGAGGCC 10235
 QY 351 AlaProSerCyThrPheThrThrGlyProLeuSerMetAlaGlyTyrluIleGlyLeuPro 370
 DB 10236 TCCCGCTGTGAGCTTCAACTCTCCAGGCGGATACAACTTCGGCACTGTGGCAGGCCG 10295
 QY 371 AlaProGlyCyAspValValLeuValProValAspGlyLeuLeuGluGluArgPheHis 390
 DB 10296 CTATGGGTATGATCAGATGACCAAGACTGAG---GACGGGAGATG-----CTCTACCGT 10346
 QY 391 GlyProHisValMetSerGlyTyrluPheAlaProGluGlnAnAlaPheAsp 410
 DB 10347 GGTCCCAACGCTCAAGAAAGATATTTGAAGGCCCTCGAGGCCCATCCCGCAGCAATC--- 10403
 QY 411 GluGluGlyTyrluTyrluCySerSerGlyAspAlaIleLeuLeuAlaAspProAlaAspGln 430
 DB 10404 GAGATGCTGCTGTGACACTGGGATATTGGCAGATC-----GACGAAGAT 10451
 QY 431 LysGlyLeuMetPheAspGlyArgIleAlaGluAspPheLeuSerSerGlyValPhe 450
 DB 10452 GGTTCCTGCTCATTAACGATCGCTCAAAAGATCATGCTCACCTTAACGCGCAAGAT 10511

```

QY 113 HisIleValGlyLeuLeuGlnProGlyLeuValPheAlaIleAspAlaIleProPheGln 132
   ||| ||| |||
   : : : : :
Db 9396 CACATGCTCATGATGCTGGGGTGCCTGTCATCATCACGGCTGGTCCCAAGAGACTGCAC 9455
QY 133 ArgAlaIleGluThrIleLeuProAspAspValProAlaIlePheThr----- 148
   ||| ||| |||
   : : : : :
Db 9456 CGATATCTCGAGGCC-----CGCGATCAAGTCCGGCCCTTGAGACGGTATCTTGATA 9509
QY 149 -----ArgGlyGluLeuAlaGlyArgArgThrValSerPheAspSerLeu 163
   ||| ||| |||
   : : : : :
Db 9510 AATCCGACATCATGATGTTGGTGCATCATGACGGTTCCTGCTCCCTGGAAACAG--- 9566
QY 164 LeuGluGlnProGlyGlyIleGlu-----AlaAspAsnAlaPheAla 178
   ||| ||| |||
   : : : : :
Db 9567 GTGGCTCAGCGTGGGCTTTCGAGAGAGATACAGACGGTCGTCGAGAGCCGATGGGCGAG 9626
QY 179 ThrGlyProAspThrIleAlaIlePheLeuPheThrSerGlySerThrIlePheProIys 198
   ||| ||| |||
   : : : : :
Db 9627 TCATGCCCGACGATGTTCGCCGCTTGATTTACCTCGGAAACACCGGCTGAGCCCAAG 9686
QY 199 AlaValProThrThrGlnArgMetLeuCysAlaAsnGlnMetLeuGlnThrPhe 218
   ||| ||| |||
   : : : : :
Db 9687 GGGGTCTGATCATCACCATCCGGCCGCTGGCTAG-----CTGCAGGCCACTT 9734
QY 219 ProValPheGlyGluGlnProProVal-----LeuValAspThrLeuProThrAsnHis 236
   ||| ||| |||
   : : : : :
Db 9735 GAGCGCTTTCGACGTCACCCGACGTCATCATCGCTGAGACTTCCTGCGCGCTCTCCAC 9794
QY 237 ThrPheGlyIleSerHisAsnIleGlyIleValLeuTyThrAsnGlyThrTyThrLeu 256
   ||| ||| |||
   : : : : :
Db 9795 GCCCTGAGTGGGAGATGTCATGCGCGTGAATTCGATTCGATTCCTCAATACCTTCGTG 9854
QY 257 AspAspGlyIlePheProThrAlaGlnGlyPheAlaGluThrLeuArgAsnLeuSerGluIle 276
   ||| ||| |||
   : : : : :
Db 9855 CCCAACCCGAAAGCATCTCG-----GCAATGTTGGCTGAGGTC 9883
QY 277 SerProThrIleTyThrLeuThrValProIysGlyTyThrGluLeuValGlyAlaLeuGlu 296
   ||| ||| |||
   : : : : :
Db 9894 CGCCCAACCTTTCGTCGTCAGTCCGAAACTGTACGAAACAGTCATG----- 9941
QY 297 ArgAspSerThrIleuArgGluArgPhePhe-----AlaArgMetIleValPhePhe 313
   ||| ||| |||
   : : : : :
Db 9942 -----AGCGTGGCCGTGAGAGGTTTCTGATTCCTCCGCAAGCTCAAAATCTTGAG 9995
QY 314 PheAla----- 315
   ||| ||| |||
   : : : : :
Db 9996 TGGTCATATACGATGGCCCGAGTGGTGGCAAGCCGACGAGGAGGACGGCCGCCAC 10055
QY 316 ---AlaAlaGlyLeuSerGlnGlyIleTyPheAspArgLeu----- 327
   ||| ||| |||
   : : : : :
Db 10056 GTATCTTGTGGAGGCGCTCATGGGGTTGCCGATGTCCTCGCTCAAGGCCATCCGAGC 10115
QY 328 ----- 329
   ||| ||| |||
   : : : : :
Db 10116 GCCATCGGTGACCCAGACTGCTGTCGAGCCGCGAGACTCCACTCCGCAAGAGACT 10175
QY 331 AlaGluGlnHisCysGlyIleArgIleArgMetMetAlaGlyLeuGlyMetThrGluThr 350
   ||| ||| |||
   : : : : :
Db 10176 GAAGAGTTCTTCCCGCTGCGGCTGTTGGTGGCCAGGAGATTCGGGCTGACGAGAGGCC 10235
QY 351 AlaProSerCysThrPheThrThrGlyProLeuSerMetAlaGlyTyThrIleGlyLeuPro 370
   ||| ||| |||
   : : : : :
Db 10236 TCCCGCTGGTGAAGCTTCACTCCAGGGGATACAAAGTTCCGCACTCGCGAGAGCCG 10295
QY 371 AlaProGlyCysGluValIleValLeuValProValAspGlyIleValLeuGluArgPheHis 390
   ||| ||| |||
   : : : : :
Db 10296 CTAGTGGGTGTGATGACACACACTGAG---GACGGGAGAGATT-----CTCTACCGT 10346
QY 391 GlyProHisValMetSerGlyTyThrArgAlaProGluGlnAsnAlaGlnAlaPheAsp 410
   ||| ||| |||
   : : : : :
Db 10347 GGTCCCAACCTCATGAAGATTTGGAAGCCCTTGAGGCCACTCGCCGACGACATC-- 10403

```

```

QY 411 GluGluGlyTyThrCysSerGlyAspAlaIleIleLeuAlaIleAspProAlaAspProGln 430
   ||| ||| |||
   : : : : :
Db 10404 GAAGATGCTGTGTTCACACTGGGGATATTGGCACATC-----GACGAAGAT 10451
QY 431 LysGlyLeuMetPheAspGlyArgIleAlaGluAspPheIleSerSerGlyValPhe 450
   ||| ||| |||
   : : : : :
Db 10452 GGTTTTCTGTCATTACCGATCCGCTCAAGACATCATCTCATCCCTTAACGGCAAGAT 10511
QY 451 ValSerValIleProLeuArgThrArgAlaValLeuGluGlyIleSerTyThrValLeuAsp 470
   ||| ||| |||
   : : : : :
Db 10512 ATTTCGCTACGCCGATCGAG-----AATTGCTCATGAAAGAC 10550
QY 471 -----ValValAlaIleProAspArgGlyCysLeuGlyLeuVal 485
   ||| ||| |||
   : : : : :
Db 10551 CCCCTTCGACACCGCGGCTGCTGGCGGATTAACCGTCCGCTCGACTCTCTGTT 10610
QY 486 PheProArgLeuLeuAspCysArgAlaLeu-----SerGlyLeu 498
   ||| ||| |||
   : : : : :
Db 10611 AAGCGTGGCTGCTCAGTGAAGAGTGGCCGAGGTTGCATATCACTCATAGACG 10670
QY 499 GlyLeuGluLaseAspAlaGluValLeuAlaSerGluProValArgAlaIlePheAla 518
   ||| ||| |||
   : : : : :
Db 10671 GCGCCGAAATGCTGCTGCTCCGAGAGCTAGCCGAGAA---ATTGTCGCGGGTGGCC 10727
QY 519 AspThrLeuIleAspArgLeuAsnArgGluAlaThrGlyAsnAlaSerArgIleMetProVal 538
   ||| ||| |||
   : : : : :
Db 10728 GAGATCACCGAATAATGCTCATCAAGACAGATTGCTGATCTGCGAGTGTGGGAC 10787
QY 539 GlyLeuLeuAspThrProProSerIleAspIleGlyGluValThrAspIleSerIle 558
   ||| ||| |||
   : : : : :
Db 10788 GAGTTC-----ACCACAGACACGAGTGTCTCATCCACCTTCAGGTA 10832
QY 559 AsnGlnArgAlaValLeuGlnThrArgSerAlaIleValAspAlaLeuTyThrArgGly-- 577
   ||| ||| |||
   : : : : :
Db 10833 CGCCGACGTAGGTGGAAGAGATTCACCGAATCTTGAAGAGATGACGCGGCTT 10892
QY 578 ---GluAspGlnSerMetLeuArgAsp 585
   ||| ||| |||
   : : : : :
Db 10893 GCGGCCGACGAAGGCTGTGAAGAAG 10920

```

RESULT 9

ACF64451
ID ACF64451 standard; DNA; 37286 BP.
XX ACF64451;
AC
XX 17-OCT-2003 (first entry)
DT
XX
XX
DE Propionibacterium acnes DNA contig sequence #17.
XX
XX Acne vulgaris; antiseborrheic; dermatological; antibacterial;
KW immunostimulant; immune response; vaccine; ds.
XX
XX Propionibacterium acnes.
OS
XX
XX
XX WO2003033515-A1.
PN
XX
XX 24-APR-2003.
PD
XX
XX 11-OCT-2002; 2002WO-US032727.
PF
XX
XX 15-OCT-2001; 2001US-00978825.
PR
XX
XX (CORI-) CORIXA CORP.
PA
XX
XX Mitcham JL, Skeiky YAM, Persing DH, Bhatta A, Maisonneuve JL,
PI Zhang Y, Wang S, Jen S, Lodes WD, Benson DR, Jones R, Carter D;
PI Barth B, Valiieve-Douglas J;
XX
XX WPI; 2003-381769/36.
DR
XX
XX
XX New Propionibacterium acnes polypeptides and polynucleotides encoding the
PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,

```

QY 385 LeuGIUGLYArpHe-----HieGIYProHieValMeSerGIYTrPrArGAlaPro 402
Db 33900 ACCGTTGGAGGGCTTACGTTCTAGTCCAGACGAGTTCGGCGGCTATCTAAATCGACCG 33959
QY 403 GIUGIAsnAlaGlnAlaPheArpGIUGLYTrYrCYSerGIYAerAlaIleLys 422
Db 33960 GAAGCCAGACGCGAGCGCTTCAAGAGGAGCGTTGGTACCGTACAGTGTAT---GTTGCG 34016
QY 423 LeuAlaerProAlaerProGlnLysGIYLeuMePheArpGIYArGIlleAlaLys 442
Db 34017 GTCGTGACACAGTGGCGGATGATCCG-----ATCGTGGGTCCGAGTGGCTTGAC 34067
QY 443 PheLysLeuSerSerGIYValPheValSerValGIYProLeuArGTrArGAlaValLeu 462
Db 34068 CTAATCAATATGGCGGATACCGGATCCGCGCAGAGAAATC-----GAGATGGCGCTG 34121
QY 463 GIUGLYGIYSerYrValLeuArpVal---ValValAlaAlaProArPArGIUCyLys 481
Db 34122 CTAGGACATCCGATGTCGGGAGCTGCGAGCTTCTCGGCTGCTGAC---GAGGATTTG 34178
QY 482 GIYLeuValPheProArGIYLeuArpCYArGAlaLeuSerGIYLeuGIYLYSGlu 501
Db 34179 GGTCAACGAGTCGTTGCATTCTGTC-----GTAGGCGCAGAG 34214
QY 502 AlSerArpAlaGIUValLeuAlaSer-----Glu 511
Db 34215 GCATTGATGCTCCGATGATGATTAACTATGTTGCCACACAGCTTTCATACATACACGG 34274
QY 512 ProValArGAlaTrPheAlaArPTrPheLysArGIYAsnArGIUAlaTrnGIYAsn 531
Db 34275 CCGCGTAACTCCGCTTCTGTCGAGCGGCTCCGCGC---AACGC-GATGGGAAAGTCTT 34330
QY 532 AlSerArGIleMeTrPValGIYLeuLeuArPTrProSerGIleArpLYSGlu 551
Db 34331 CAAAGAGCACTGCTCTTCGAGGCGTACTGCGCTACGACGAGTACTGATTATGCGGCG 34390
QY 552 ValThrArpLYSGLYSerIleAsnGIYAlaValLeuGIYTrP----- 566
Db 34391 ACCGCATCGTTCCGACCGGACCGGGGCGGAGTGTGTGTTGGTCCGCGCCACTGT 34450
QY 567 ArgSerAlaLYValArPAlaLeuTrP-----ArgGIYGIUArpGlnSer 581
Db 34451 AGAGCTTCAAGAGCAGCAGCATGTGACACCTAGTGTGTCGAGGGGTCCGGTAGAGGT 34510
QY 582 MetLeuArGArp 585
Db 34511 TTTTCAGCGGAC 34522

```

RESULT 8
AAS59522 standard; DNA; 37286 BP.

AAS59522:

13-FEB-2002 (first entry)

Propionibacterium acne immunogenic protein encoding DNA #17.

SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
dermatological; osteoporotic; neuroprotectant; ds.

Propionibacterium acne.

MO200181581-A2.

01-NOV-2001.

20-APR-2001; 2001MO-US012865.

21-APR-2000; 2000US-0199047P.

```

PR 02-JUN-2000; 2000US-0208841P.
PR 07-JUL-2000; 2000US-0216747P.
PA (CORI-). CORIXA CORP.
XX Skelky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX WPI; 2001-616774/71.
DR
PT Propionibacterium acne polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris.
PS Claim 1; SEQ ID NO 17; 1069pp; English.
XX
CC Sequences AAS59506-AAS59804 represent DNA molecules encoding
CC Propionibacterium acne immunogenic polypeptides. The proteins and their
CC associated DNA sequences are used in the treatment, prevention and
CC diagnosis of medical conditions caused by P. acne. The disorders include
CC SAPHO syndrome (synovitis, acne, pustulosis, hypertosis and
CC osteomyelitis), uveitis and endophthalmitis. P. acne is also involved in
CC infections of bone, joints and the central nervous system, however it is
CC particularly involved in the inflammatory lesions associated with acne
CC vulgaris. A method for detecting the presence or absence of P. acne in a
CC patient comprises contacting a sample with a binding agent that binds to
CC the proteins of the invention and determining the amount of bound protein
CC in the sample. The polypeptides may be used as antigens in the production
CC of antibodies specific for P. acne proteins. These antibodies can be
CC used to downregulate expression and activity of P. acne polypeptides and
CC therefore treat P. acne infections. The antibodies may also be used as
CC diagnostic agents for determining P. acne presence, for example, by
CC enzyme linked immunosorbent assay (ELISA). This sequence encodes the
CC polypeptides shown in AAU44158-AAU44568 and AAU67488-AAU67490. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 37286 BP; 7034 A; 10065 C; 11922 G; 8264 T; 0 U; 1 Other;

```

Alignment Scores:

Pred. No.:	Length:	Score:
1,52e-15	37286	301.50
Percent Similarity:	Matches:	154
Best Local Similarity:	Conservative:	108
Query Match:	Mismatches:	264
DB:	Gaps:	144
		29

US-09-750-986D-30 (1-589) x AAS59522 (1-37286)

```

QY 9 ProPheProGIYArG-----IleLeuGIUArGLeuGIUTrPrAlaLysTrP----- 24
Db 9066 CCGTACCGGGGAGAACTTCAATGAGCCACCTGCGCCATGTTCCGCGCCAGCGTT 9125
QY 25 -----ArgProGIUTrnTrCYValAlaAlaArGAla---AlaAsnGIY 38
Db 9126 GCCAACCAAGGGTTCGCTCC-----GCCACTGGGTGCTGACGGCGGT 9170
QY 39 GIUTrArGArGIleSerYrAlaGluMePheHieAsnValArGAlaIleAlaGlnSer 58
Db 9171 CAGTGGATTATCCGACATATGCGAGACTGGGCGTGGCTATACCGGGCTCGCGCGGCC 9230
QY 59 LeuLeuProTYrGIYLeuSerAlaGIUArGProLeu-----LeuIle 72
Db 9231 TTGTGCACTCCCGGGGTGCTCACCGAGATGGCTCAGCGCGGGAGTGTATTTCTGTG 9290
QY 73 ValSerGIYAsnArPLeuGIUTrnIleValPheGIYAlaMeTrYrAlaGIYLeu 92
Db 9291 TTGCGCGGATATGCGCGAGATGATCGAGCGTATCGCGGCGATGAGATGGCGGTG 9350
QY 93 ProTYrCYArProValSerProAlaTYrSerLeuLeuSerGIleArpLeuAlaLysLeuArG 112
Db 9351 -----ATCCCGGTG---CCGATTACCGGACGTGACCCCTGAT-----CAGATTGTT 9395

```

Db 18317 CACGCCATCGTTGGACCGGACGGGGCCGGATGTGTGGTCGCCGCCCACTG 18376
|||||
Qy 567 -ArgSerAlaIysValAspAlaLeuTyr-----ArgGlyIuAepGlnSe 581
|||||
Db 18377 TAGAGCTTCAAGACCCAGCGCAATGTGGACCGTACTGTGTGTGCGAGGGGTCCGGGTAGAG 18436
Qy 581 MetLeuArgAsp 585
|||||
Db 18437 TTTTTCAGCGGAC 18449
RESULT 7
ADB74389
ID ADB74389 standard; DNA, 35961 BP.
XX ADB74389;
XX
XX 04-DEC-2003 (first entry)
XX
XX Mycobacterium leprae DNA #23.
XX
XX Non-naturally occurring peptide; anion pump protein; tuberculosis;
KM hyperemotivity reaction; tuberculostatic; gene; ds.
XX Mycobacterium leprae.
XX
XX US6583266-B1.
XX
XX 24-JUN-2003.
XX
XX 16-SEP-1994; 94US-00311731.
XX
XX 19-AUG-1993; 93US-00109181.
XX 22-OCT-1993; 93US-00142558.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Smith DR, Mao J;
XX
XX WPI; 2003-656441/62.
XX
XX New Mycobacterium tuberculosis anion pump peptide useful for as
PT tuberculosis vaccine and diagnosis of tuberculosis infection.
XX
XX
XX Disclosure; SEQ ID NO 138; 26pp; English.
XX
XX The invention relates to a non-naturally occurring peptide of
CC Mycobacterium tuberculosis comprising an amino acid sequence
CC corresponding to an anion pump protein. The invention also relates to a
CC non-naturally occurring nucleic acid corresponding to a DNA sequence of
CC Mycobacterium tuberculosis or Mycobacterium leprae. The new peptide is
CC useful as a vaccine against Mycobacterium tuberculosis or Mycobacterium
CC leprae or for screening for new tuberculosis drugs. Purified proteins
CC derived from the sequences of the invention may elicit a specific immune
CC response. The peptide may also be used to detect hypersensitivity
CC reactions of individuals exposed to Mycobacterium tuberculosis or
CC Mycobacterium leprae. The proteins and peptides may be affixed to solid
CC supports to detect antibodies typical of hypersensitivity reactions, from
CC a patient's sera. This sequence represents Mycobacterium leprae DNA of
CC the invention. Note: The sequence data for this patent did not form part
CC of the printed specification but was obtained in electronic format
CC directly from USPTO at seqdata.uspto.gov/sequence.html.
XX
XX
SQ Sequence 35961 BP; 6748 A; 9918 C; 11140 G; 8155 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1.18e-15 Length: 35961
Score: 302.50 Matches: 150
Percent Similarity: 37.50% Conservative: 69
Best Local Similarity: 25.68% Mismatches: 234
Query Match: 9.82% Indels: 132
DB: 10 Gaps: 27

US-09-750-986D-30 (1-589) x ADB74389 (1-35961)
Qy 44 SerTyrAlaIuMetPheHisAsnValArgAlaIleAlaIuInsLeuLeuProTyrGly 63
|||||
Db 33039 TCGGTGGCTGACGCGTGTGCGAGGTGCACGCCCTGTATACGGGTGCGCACGCCG----- 33092
Qy 64 LeuSerAlaGluArgProLeuLeuIleValSerGlyAsnAspLeuGluHisLeuGlnLeu 83
|||||
Db 33093 ---ACGCGATCGACTGTGCTAGCGATTACT----- 33119
Qy 84 AlaPheGlyAlaMetTyrAlaGlyIleProTyrCysProValSerProAlaTyrSerLeu 103
|||||
Db 33120 ---GGTGCCTGATGCCGGGTGTCGGGTGTCCGGGTGCCCGGATATCGGTGTG 33173
Qy 104 LeuSerGlnAspLeuAlaLysLeuArgHisIleVal-----GlyLeu 117
|||||
Db 33174 GTCGACGC-----CGGCATATGCTACACGATTCGTGGGGCGAGCGCTCG 33218
Qy 118 LeuGlnProGlyLeuValPheAlaAlaAspAlaAlaProPheGlnArgAlaIleGlyThr 137
|||||
Db 33219 CTGGGTCCGGGGCC-----AGTCCGACTCCGCG----- 33248
Qy 138 IleLeuProAspAspValProAlaIlePheThrArgGlyGlyLeuAlaGlyArgThr 157
|||||
Db 33249 ---CGGACGGGTGGCCGCATATC-----CCG 33272
Qy 158 ValSerPheAspSerLeuLeuGlnGlnProGlyGlyIleGluAlaAspAsnAlaPheAla 177
|||||
Db 33273 GTTCAACTAGATGT-----CGCTCTGGAACCGTATCT 33308
Qy 178 AlaThrGlyProAspThrIleAlaLysPheLeuPheThrSerGlySerThrLysLeuPro 197
|||||
Db 33309 GAGCGGTCACTGATGACCGCGAGTGTGATGATCAACCTCGGGTACTAGTGGCCCCCT 33368
Qy 198 LysAlaValProThrThrGlnArgMetLeuCysAlaIasnGlnMetLeuLeuGlnThr 217
|||||
Db 33369 AAGGGGTGATGTTGACCGCGACGCGATGCGTATTGACGCGCTTGGCCAGGCC 33428
Qy 218 PheProValPheGlyGluGlnProProValIleuValAspThrLeuProTyrPheHisIleThr 237
|||||
Db 33429 TGGCAGTGCAGCGCCCATGAC-----GTTTGGTACATAGATTGCCGCTTTCAAGTGG 33482
Qy 238 PheGly-----GlySerHisAsnIleGlyIleValLeuTyrAsnGly 251
|||||
Db 33483 CACGGCTTGATCGGTTGCTTGATGCTTCGATG----- 33521
Qy 252 GlyThrTyrTyrLeuAspAspGlyLysProThrAlaIleGlnIlePheAlaGluThrLeuArg 271
|||||
Db 33522 GGAATGCTTGTGTACACACCGGAAACCCACTCCGACCGCTATGCTCAGGCT----- 33575
Qy 272 AsnLeuSerGlnLysSerProThrAlaTyrLeuThrValProLysGlyTyrGlnGluLeu 291
|||||
Db 33576 ---TGCTTTAAGCGCGGGGATCTCTGTATTTGGCGCTTCCACGATATGTCGGGCTG 33632
Qy 292 ValGlyAlaLeuGluArgAspSerThrLeuArgGluArgPhePheAlaArgMetLysLeu 311
|||||
Db 33633 GTGGCA-----GACGAGGACAGT---GCCGGCGGTTAAGACCGCGCGGCTA 33677
Qy 312 PhePhePheAlaAlaAlaGlyLysSerGlnGlyIleThrAspArgLeuAspArgValAla 331
|||||
Db 33678 CTAGTGTCTGGAGGTGCTGCTATGCGCGTTCGCGTATTTGATGCGGTAGCGCACTTCA 33737
Qy 332 GlnGlnHisCysGlyGluArgGlyLeuArgMetMetAlaGlyLeuGlyMetThrGluThrAla 351
|||||
Db 33738 GGTTCATCGGCCCATCGAACGG-----TATGGACAGCAGGAATTCATG 33779
Qy 352 ProSerCysThrPheThrThrGlyProLeuSerMetAlaGlyTyrIleGlyLeuProAla 371
|||||
Db 33780 ATCACCCTGTGCACGCTTGGGAGACGAGACGCGCGCTGCGGTGGGATTTGCCGCTG 33839
Qy 372 ProGlyCysGluValLysLeu-----ValProValAspGlyLys 384
|||||
Db 33840 GCGGGCGGTGCAGACCAAGCTGTGATGATCGGAGGTCCAGTGCCTTACGACGGGGA 33899

DR ADB74323, ADB74324, ADB74325, ADB74326, ADB74327, ADB74328, ADB74329,
 DR ADB74330, ADB74331, ADB74332.
 XX New Mycobacterium tuberculosis anion pump peptide useful for as
 PT tuberculosis vaccine and diagnosis of tuberculosis infection.
 PS Disclosure; SEQ ID NO 24; 26pp; English.
 XX

CC The invention relates to a non-naturally occurring peptide of
 CC Mycobacterium tuberculosis comprising an amino acid sequence
 CC corresponding to an anion pump protein. The invention also relates to a
 CC non-naturally occurring nucleic acid corresponding to a DNA sequence of
 CC Mycobacterium tuberculosis or Mycobacterium leprae. The new peptide is
 CC useful as a vaccine against Mycobacterium tuberculosis or Mycobacterium
 CC leprae or for screening for new tuberculosis drugs. Purified proteins
 CC derived from the sequences of the invention may elicit a specific immune
 CC response. The peptide may also be used to detect hypersensitivity
 CC reactions of individuals exposed to Mycobacterium tuberculosis or
 CC Mycobacterium leprae. The proteins and peptides may be affixed to solid
 CC supports to detect antibodies typical of hypersensitivity reactions, from
 CC a patient's sera. This sequence represents Mycobacterium leprae DNA of
 CC the invention. Note: The sequence data for this patent did not form part
 CC of the printed specification but was obtained in electronic format
 CC directly from USPTO at seqdata.uspto.gov/sequence.html.
 CC

XX Sequence 38494 BP; 7322 A; 10677 C; 12008 G; 8487 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 3,31e-16 Length: 38494
 Score: 309.00 Matches: 149
 Percent Similarity: 37.78% Conservative: 72
 Best Local Similarity: 25.47% Mismatches: 231
 Query Match: 10.03% Indels: 133
 DB: 10 Gaps: 27

US-09-750-986d-30 (1-589) x ADB74275 (1-38494)

QY 44 SerTyrAlaGluMetPheHisAsnValAlaAlaIleLeuGlnSerLeuLeuProTyrGly 63
 DB 16967 TCGGTGCTAGAGCTGTGCGAGGTGACGCGCTGTACGGTCTCGCCAGCCGCG- 17020
 QY 64 LeuSerAlaGluArgProLeuLeuIleValSerGlyAsnAspLeuGlnIleLeu 83
 DB 17021 -ACGGCATCGACTGTGATGAGGATCTACT- 17047
 QY 84 AlaPheGlyAlaMetTyrAlaGlyIleProTyrCysProValSerProAlaTyrSerLeu 103
 DB 17048 -GGTTGCCTGATTGCGCGGTGTCGCGGTGTCGCGGTGTCGCGGATATCGGTGTG 17101
 QY 104 LeuSerGlnAspLeuAlaValSerArgHisIleVal- 17104
 DB 17102 GTGGAACGC- 17146
 QY 118 LeuGlnProGlyLeuValPheAlaAlaAspAlaAlaProPheGlnArgAlaIleGluThr 137
 DB 17147 CTGGGTCCGGGCGCC-AGTGGGACTCGGCG- 17176
 QY 138 IleLeuProAspAspValProAlaIlePheThrArgGlyLeuLeuAlaGlyArgGlyThr 157
 DB 17177 -CCGGACGGGTGGCGCATATC- 17200
 QY 158 ValSerPheAspSerLeuLeuGlnProGlyGlyIleGlnAlaAspAsnAlaPheAla 177
 DB 17201 GTTCACTAGATGCT- 17236
 QY 178 AlaThrGlyProAspThrIleAlaValPheLeuPheThrSerGlySerThrIleValPro 197
 DB 17237 GACCGGTCACTGATGACCGGATGTGATCATACCTCGGATCTACTGAGCCCTT 17296
 QY 198 LysAlaValProThrThrGlnArgMetLeuCysAlaAsnGlnGlnMetLeuGlnThr 217
 DB 17297 AAGGGGTACTGTGGTGGCGGACGATGCGCTGTGATTGGACGCGCTTGGCCAGGCC 17356

QY 218 PheProValPheGlyGluGluProProValIleAspTrpLeuProTyrPheAsnIleThr 237
 DB 17357 TGGCACTGACAGCCCATGAC- 17410
 QY 238 PheGly- 17411
 DB 17411 CACGGCTTGTACTCGGTTGCTTGTGATCGCTCGGATT- 17449
 QY 252 GlyThrTyrTyrLeuAspAspGlyIleProThrAlaGlnGlyPheAlaGluThrLeuArg 271
 DB 17450 GGAATTCGCTTTGTACACACCGGAAACCCACACCCAGCCCTATCTCAGGCT- 17503
 QY 272 AsnLeuSerGluIleSerProThrAlaTyrLeuThrValProGlyGlyTrpGluGluLeu 291
 DB 17504 -TGCCTGACCCGGGGATCTCTGATTTTGGCGTTCACCGGTATGTTGCGGCTG 17560
 QY 292 ValGlyAlaLeuGluArgAspSerThrLeuArgGluArgPhePheAlaArgMetLeu 311
 DB 17561 GTGGCA- 17605
 QY 312 PhePhePheAlaAlaAlaGlyLeuSerGlnGlyIleTrpAspArgLeuAspArgValAla 331
 DB 17606 CTAGTGTCTGGAGTGCCTCATTTGCGGTTCCGTTTGTGATCGGTACCGCACCTACA 17665
 QY 332 GluGlnHisCysGlyGluArgIleArgMetMetAlaGlyLeuGlyMetThrGluThrAla 351
 DB 17666 GGTCAATCGGCCCATCGAACG- 17707
 QY 352 ProSerCysThrPheThrThrGlyProLeuSerMetAlaGlyTyrIleGlyLeuProAla 371
 DB 17708 ATCACTGTGCGACGTGGCGGACGCGTGAACGCGCGGTGGGTGGATTTGCGCGCTG 17767
 QY 372 ProGlyCysGluValValLeu- 384
 DB 17768 GCGGGGTGTCAGACACAGCTGTGATGATGATCGGAGTCCAGTTCAGTTCACGCGGGA 17827
 QY 385 LeuGluGlyArgPhe- 402
 DB 17828 ACCGTGGAGGCTTACAGTTGTTAGTCCGACGATTTGGCGGCTATCTAAATCAACCG 17887
 QY 403 GluGlnAsnAlaGlnAlaPheAspGluGluGlyTyrTyrCysSerGlyAspAlaIleVal 422
 DB 17888 GAAGCAGACAGCGAGCGTTTCGACGAGACGAGTGTGATCGTACAGATAT- 17944
 QY 423 LeuAlaAspProAlaAspProGlnIleGlyLeuMetPheAspGlyArgIleAlaGluAsp 442
 DB 17945 GTCTGGACAGTGGCGGATGATCGC- 17995
 QY 443 PheLeuLeuSerSerGlyValPheValSerValGlyProLeuArgThrArgAlaValLeu 462
 DB 17996 CTATCAAAATTGGGCGGATACCGGATCGCGCGCAGGAAATC- 18049
 QY 463 GluGlyGlySerTyrValLeuAspVal- 481
 DB 18050 CTAGACATCCGATGTGCGGAGGTGCGCATGTGTGCGCTGCTGAC- 18106
 QY 482 GlyLeuLeuValPheProArgLeuLeuAspCysArgAlaLeuSerGlyLeuGlyGly 501
 DB 18107 GGTCAACGATCTGTCATCTTC- 18142
 QY 502 AlaSerAspAlaGluValLeuAlaSer- 511
 DB 18143 GATTTAGTCCGATAGTTATTAATCTATGTTCCCAACAGCTTTCATACATAGCGG 18202
 QY 512 Pro-ValArgAlaTrpPheAlaAspTrpLeuValArgLeuAsnArgGluAlaThrGlyAs 531
 DB 18203 CCGCGTGAAGTCC- 18256
 QY 531 nAlaSerArgIleMetTrpValGlyLeuLeuAspThrProProSerIleAspValGly 551
 DB 18257 TCAGAAAGAGATGCTCTTTCGAGGCTGACTGCGCTACGACGAGTATTAATGCGGG 18316
 QY 551 uValThrAspLysGlySerIleAsnGlnArgAlaValLeuGlnIleTrp- 566

Db 54975 TCGGCGCCAGGCGCTGATCGCGCTCGCATCGCAAGGCGACGGCGTGACCATCTT 55034
 Qy 74 SerGlyAsnAspLeuGluHisLeuGluLeuAlaPheGlyAlaMetYrAlaGlyLeuPro 93
 Db 55035 TCCTCCACCCGCTCTAGATAGATGGGCGATTCTAGATTTCGACTGGCGCGGCGGCGCGGTA 55094
 Qy 94 TyrCysProVal-----SerProAlaTyrSerLeuLeuSerGlnAsp 107
 Db 55095 AGCGTCGCGATCTATGACACCGAATCGCGCCCGACGCGCATCATGAAGATTTCG 55154
 Qy 108 LeuAlaAlaSerLeuArgHisLeuValGlyLeuLeuGlnProGlyLeuValPheAlaAlaAsp 127
 Db 55155 GTGGTCAAG-----CTGGCTTTGGGACGCTAGAGCGTGACC 55181
 Qy 128 AlaAlaProPheGlnArg-----AlaIleGluThrIle 138
 Db 55182 CGGAGAGGCTTTCAGCCGGCTCGACTCGGTCAAGACCACTGCCCTGCTTGAACAATTT 55241
 Qy 139 LeuProAspAspValProAlaIlePheThrArgGlyGlyLeuAlaGlyArgArg---Thr 157
 Db 55242 CTCATGATCGAGGCGACGCGCTG-----GGCGCTTTGGAGGCGCTAGAGCGTGACC 55292
 Qy 158 ValSerPheAspSerLeuLeuGlnGlnProGlyGlyIleGluAlaAspAsnAlaPheAla 177
 Db 55293 GTTTCGACGAGAACTGACAGAGCGGTGGCCACCGTGGCGACAGAGAC----- 55343
 Qy 178 AlaThrGlyProAspThrIleAlaIlePheLeuPheThrSerGlySerThrIleLeuPro 197
 Db 55344 -----CTGGCCACCATGCTCTACACTTCGGGCTCCACCGGCAACCCG 55385
 Qy 198 LysAlaValProThrThrGlnArg-----MetLeuCyAlaAsnGlnGln 212
 Db 55386 AAGGGTCCGAACTGACCCACAGAACTTGCTGTCATACACATCTCCGCTCGCAGGCG 55445
 Qy 213 MetLeuLeuGlnThrPheProValPheGlyGlyGluProProValLeuValAspTrpLeu 232
 Db 55446 TTG-----CATGAAGTGGTCTTGACGACCATCGCGCTGCTCTCTCTG 55493
 Qy 233 ProTrpAsnHisThrPheGlyGlySerHisAsnIleGlyLeuValLeuYrAsnGlyGly 252
 Db 55494 CCGCTGGCCCATGCTTGGCGCATTCATCCAGTACGCTCCATCGCTCGCATGACGCG 55553
 Qy 253 ThrTyr---TyrLeuAspAspGlyLysProThrAlaGlnGlyPheAlaGluThrLeuArg 271
 Db 55554 GTGGTGGTATCTGCGCGACACCAA-----ACCGTCTGCGG 55592
 Qy 272 AsnLeuSerGlnIleSerProThrAlaTyrLeuThrValProLysGlyTrpGluGluLeu 291
 Db 55593 GACCTGGTTCGTTTGAACGACCTTACCTGCTGGCGCTGCGCGCTTGAAGAGTA 55652
 Qy 292 ValGlyValLeuGluArgAspSerThrLeuArgGluArgPhePheAlaArgMetLysLeu 311
 Db 55653 TACACGCCGCCCTCGCACAGGCGCGT-----GCCGCTGAAAGGCG 55694
 Qy 312 PhePhePheAlaAlaAlaGlyLeuSerGlnGlyIleTrpAspArgLeuAspArgValAla 331
 Db 55695 CGCGTTTGTCAAGCGCGCGAGCGCGCGCTGCGAGCGCAAGAGAGCGCGCG 55754
 Qy 332 GluGlnHis----- 334
 Db 55755 GACGACGACACGTTTCCGAATCGCGACGCGCCAAAGTACGAAGACGCTCGTCTACCG 55814
 Qy 335 -----CysGlyGlyLysArg----- 338
 Db 55815 ACGGTGCGCGGTGCTGCGGCCGCAAGATCAAGTACGTGGCTGCGGCGCGCACCTG 55874
 Qy 339 -----IleArgMetMetAlaGlyLeuGly 346
 Db 55875 TCGCTGATCTGGGCAATTTCTACACGCGATCGGATCGCATGATTCAGGGTTATGCG 55934
 Qy 347 MetThrGluThrAlaProSerCysThrPheThrGlyProLeuSerMetAlaGlyTyr 366
 Db 55935 ATGACGGAACCGCGCGCTTCCGCGCACGCGCGCTGACCGACATGTGATCGGCGACC 55994

Qy 367 IleGlyLeuProAlaProGlyCysGlyValLysLeuValProValAspGlyLysLeuGlu 386
 Db 55995 GTAGGCGAGCCCGCGCGCTCGCTCGATTTCG---ATTCCGACGAAAGCGAATCTCAG 56051
 Qy 387 GlyArgPheHisGlyProHisValMetSerGlyTyrTrpArgAlaProGluGlnAsnAla 406
 Db 56052 GTCAG-----GGCCGAACGTTTCCGCGGTACCAATCTGCGGGAAGACGCGG 56105
 Qy 407 GluAlaPheAspGluGluGlyTyrTyrCysSerGlyAspAlaIleLysLeuAlaAspPro 426
 Db 56106 GAGCGTTTACCGCCGACGCTGCTGACAGACCGGCACTTGGCCGAGATC----- 56156
 Qy 427 AlaAspProGlnLysGlyLeuMetPheAspGlyArgIleAlaGluAspPheLysLeuSer 446
 Db 56157 ---GATGATGAGGCGCAATCGTATCATCACCGCGCATCAAGACATTAATTATCTGCC 56213
 Qy 447 SerGlyValPheValSerValGlyProLeuArgThrArgAlaValLeuGluGlyLysSer 466
 Db 56214 GCGCGCAAGAACGTTCCCATTCCTCGCTGAGAGAGAG-----ATCGCCAAAGTGGCCG 56267
 Qy 467 TyrValLeuAspValValValAlaAlaProAspArgGluCysLeuGlyLeuValPhe 486
 Db 56268 ATCGTGAGCATGCGCGTGGTGGGCGACCAAGGCGCTTCATCGCGCGTTGATG--- 56324
 Qy 487 ProArgLeuLeuAspCysArgAlaLeuSer-----GlyLeuGlyLys 500
 Db 56325 ---ACGTTGATCCGAGCGTGGTGGCTGATTTGGCTGCCGCGACGCGCTGTCCAGC 56378
 Qy 501 GluAlaSerAspAlaGluValLeuAlaSerGlyPro-----ValArgAlaTrpPheAl 518
 Db 56379 GAG---ACGCGGTGACCGCGCTCGCCACGAAACCGCGGTGCGCGAGATTTCAGCAG 56435
 Qy 518 AspTrpLeuLysArgLeuAsnArgGluAlaThrGlyAsn-----AlaSerArgIleLeu 536
 Db 56436 TACGTGACAGGCGAATGACAGCGTCTCGCGCGCAATCGTACGCAAGTTGCGCGCTG 56495
 Qy 536 Trp 537
 Db 56496 CTGG 56499
 RESULT 6
 ADB74275
 ID ADB74275 standard; DNA; 38494 BP.
 XX
 AC ADB74275;
 XX
 DT 04-DEC-2003 (first entry)
 XX
 DE Mycobacterium leprae DNA #2.
 XX
 KW Non-naturally occurring peptide; anion pump protein; tuberculosis;
 KM hypersensitivity reaction; tuberculosis; gene; ds.
 XX
 OS Mycobacterium leprae.
 XX
 PN US6583266-B1.
 XX
 PD 24-JUN-2003.
 XX
 PF 16-SEP-1994; 94US-00311731.
 XX
 PR 19-AUG-1993; 93US-00109181.
 XX
 PR 22-OCT-1993; 93US-00142558.
 XX
 PA (GENO-) GENOME THERAPEUTICS CORP.
 XX
 PI Smith DR, Mao J;
 XX
 DR WPI; 2003-656441/62.
 DR P-PSDB; ADB74303, ADB74304, ADB74305, ADB74306, ADB74307, ADB74308,
 DR ADB74309, ADB74310, ADB74311, ADB74312, ADB74313, ADB74314, ADB74315,
 DR ADB74316, ADB74317, ADB74318, ADB74319, ADB74320, ADB74321, ADB74322,

	CC	the organism accumulates the intermediates conferyl alcohol (CAI) ,
	CC	conferyl aldehyde (CA2), FA, vanillin (V) and/or vanillic acid (VA).
	CC	INDEPENDENT CLAIHS are also included for the following: (a) gene
	CC	structures in which sequences that encode one or more of the enzymes CAI-
	CC	dehydrogenase, CA2-dehydrogenase, PA-CoA-synthetase, enoyl-CoA hydratase-
	CC	aldehyde, beta-ketothiolase, V-dihydroxyketone and VA-demethylase have been
	CC	altered and/or inactivated; (b) vector containing at least one gene
	CC	structure of (a); (c) biotechnical production of organic compounds
	CC	(particularly alcohols, aldehydes or organic acids) using (A); and (d)
	CC	production method of (A). (A) are used to prepare organic compounds,
	CC	preferably alcohols, aldehydes and organic acids, especially CAI, CA2,
	CC	FA, V and/or VA, most particularly V, a well-known natural flavouring.
	CC	The inexpensive starting material eugenol can be converted to V in a
	CC	single step. This sequence encodes the Pseudomonas sp. HR19 ferulic acid
	CC	-CoA-synthetase, fcs-delta which is described in the method of the
	CC	invention
	XX	
SO		Sequence 1203 BP; 217 A; 337 C; 357 G; 292 T; 0 U; 0 Other;
	Alignment Scores:	
	Pred. No.:	1,23e-53 Length: 1203
	Score:	698.50 Matches: 139
	Percent Similarity:	94.04% Conservative: 3
	Best Local Similarity:	92.05% Mismatches: 8
	Query Match:	22.66% Indels: 1
	DB:	Gaps: 3
US-09-750-986D-30 (1-589) x AAA39722 (1-1203)		
OY	1	MctArgSerLeuGluAlaLeuLeuProPheProGlyArgGlleuGluArgLeuGluHis 20
Db	342	ATGCCTTCCTCGAGCGCTTCTCCCTCCGGGCGAAATTGTGAAGCCTCTCGACAT 4010
OY	21	TTPalaLyThArpProGluGlnThrCySValaLaAlaARglaLaIsengLyGIUTrp 40
Db	402	TGGCGTTAAGACCCTGCCAACAACACTCGTGCTGCCAGGCGGCAATGGGAAATGG 4610
OY	41	ARGarGlleSerTYraGluWecPheHISaenValARgaLlaIleaInserLeuLeu 60
Db	462	CGTCGTATCACGCTACGCGGAAATGTTCCAACACGTCCGGCCCATCCACAGACTTGCTT 5210
OY	61	ProTYrGlyLeuSerAlaGluArgProLeuLeuIleValSerCylAsnAspLeuGluHis 80
Db	522	CCTTAACGACTATCGGCAAGCGTCCGCTTATGCTCTCGAAATGACCTGGAACAT 5810
OY	81	LeuGlnLeuAlaPheGlyAlaMetTYraGlyIleProTYrCySProValSerProAla 100
Db	582	CTTCACGTGGCAITTTGGGCTATGTAATGGGGCATTTCCATATGGCCCCGGGTCTCTGCT 6410
OY	101	TYrSerLeuLeuSerGlnAspLeuAlaLYLeuArgHisIleValGlyLeuLeuGlnPro 120
Db	642	TATTCACGTGCTGCGCAAGATTGGGGAGAAGCTCGTCACATCGTAGGTCCTTGCACACG 7010
OY	121	GlyLeuValPheAlaAlaAspAlaAlaProPheGlnARglaIleGluThrIleLeuPro 140
Db	702	GGACGCGCTTTTGTCGGCCGATGCACAGACCTTTTCCACGGCGCTTTTGCAGATGGCGGTG 7610
OY	141	AspAspValProAlaIlePheThrArpGlyGly 151
Db	762	GCGAAGATTGATGCCCTGTAT---CGTGGTGA 791
RESULT 5		
ABQ81844		
ID	ABQ81844	standard; DNA; 349980 BP.
XX	ABQ81844;	
DT	19-NOV-2002	(first entry)
DE	Bifidobacterium longum NCC2705 related nucleotide sequence SEQ ID:1100.	
KM	Bifidobacterium longum NCC2705; Bifidobacterium; bacterial.	
KW	antidiarrhetic; antibacterial; inhibitor of salmonella; detection;	

[illegible]

```

Db      702  GGACTGCTCTTGGCCGATGACGACCTTCCAG----- 737
Qy      141  AspaSpValProAlaIlePheThrArgGlyGluLeuValaGlyArgThrValSerPhe 160
Db      738  -----GGGAGAGG----- 746
Qy      161  AspSerLeuLeuGluGlnProGlyGlyIleGluValaAspAlaIlePheAlaThrGly 180
Db      746  ----- 746
Qy      181  ProAspThrIleAlaLysPheLeuPheThrSerGlySerThrIleGluProGlyAlaVal 200
Db      747  -----CGGTTTGCCTATTGGCCGATGACATATAAACTGTGTAATTCATT 791
Qy      201  ProThrThrGlnArgMetLeuGlyAlaAsnGlnGlnMetLeuLeuGlnThrPheProVal 220
Db      792  AAGCATTTCTGCCGACATGGAAGCCATCACAACGGCATGATGAA-CTGAAATGCCACGCG 850
Qy      221  PheGlyGluGluProProValLeuValaAspTrpLeuProTrpAsnIleThrPhe----- 238
Db      851  GCATCAGCACCTTGCCCTTGCGTATAATATTGGCCATGAGCAGCACCCGTGGAAACG 910
Qy      239  ---GlyGlySerHisAsnIleGlyIleValLeuTyraSerGlyThrTyrrTyrrLeuAsp 257
Db      911  GATGAAAGCAGAACCCAGTTCATAGCCTGTTCGGTTCGTAACTGTA----- 961
Qy      258  AspGlyLysProThrAlaGlnGlyPheAlaGlnThrLeuArgAsnLeuSerGluIleSer 277
Db      962  -----ATGCAGTAGCGGTATGCCCTACCGCATCTGGTCCAGAACCTTGACCGAACCGAG 1015
Qy      278  Pro-----ThrAlaTyrrLeuThrVal 284
Db      1016  CGGTGTAAAGCGCGCAGTGGCGGTTTTCATGGCTGTATGACCTGTTTTTTGTAACGTC 1075
Qy      285  ProLysGlyTrpGluGluLeuValaGlyAlaLeuGluArgAspSerThrLeu----- 301
Db      1076  TATGGCTCGGGACATCAAGCAGCAGCGCGTTACGCGGTGATGATGATGATGATGATG 1135
Qy      302  ---ArgGluArgPhePheAlaArgMetLysLeuPhePhePheAlaAlaIleGlyLeuSer 320
Db      1136  GAGCAGCAGACGATGTTCGCGACGACGACGATGTTCACGACGACGAGCGTCCCTAAAA 1195
Qy      321  GlnGly-----IleTrpAspArgLeuAspArgValaAla----- 331
Db      1196  CAAAGTAGAGTGCTCAAGATATGGCATTCATTCGCACATGATGAGCTGGCCCTGACCAAG 1255
Qy      332  GluGlnHisGlyGlyGluArgIleArgMetMetAlaGlyLeuGlyMetThrGluThrAla 351
Db      1256  TCAAAATCCATGCGGG-----CTGCTCTTGTATCTTTTCGTCGTGAGTCGAGACCTAG 1309
Qy      352  Pro-----SerCysThrPheThrThrGlyProLeuSerMetAlaGly 365
Db      1310  CCACCTTCTCCCAACATCAGCCGACCTCCGATTAACCTGGGGAAGCTCTCCGTAAAGA 1369
Qy      366  TyrIleGlyLeuProAlaProGlyCysGluValaLysLeuValProValaAspGlyLysLeu 385
Db      1370  CATTTCATCGGCGTTGCGCTTCGACCAAGAGCGTTGTTGGCGCTC---TCCGCGCTT 1426
Qy      386  GluGlyArgPheHisGlyProHisValaMetSerGlyTyrrTrpArgAlaProGluGlnAsn 405
Db      1427  ACG----- 1429
Qy      406  AlaGlnAlaPheAspGluGluGlyTyrrTyrrCysSerGlyAspAlaIleLysLeuAlaAsp 425
Db      1430  -----TTCGCCACAGT----- 1441
Qy      426  ProAlaAspProGlnLysGlyLeuMetPheAspGlyArgIleAlaGluAspPheLysLeu 445
Db      1442  -----TTG 1444
Qy      446  SerSerGlyValaPheValSerValaGlyProLeuArgThrArgAlaValaLeuGluGlyGly 465

```

```

Db      1445  AGCAGCCGCGTAGATCTATATCTATGATCTTCGACGTCTCCGCGAGACACCGAGGC 1504
Qy      466  SerTyrrValLeuAspValaValaAlaAlaProAspArgGluCysLeuGlyLeuLeuVal 485
Db      1505  AGG-----GCATTGCCACCGCGCTTCATCAATCTCTCAAGCATG 1543
Qy      486  PheProArgLeuLeu-----AspCysArgAlaLeuSerGlyLysGlyLys 500
Db      1544  AGGCCAAGCGCGTGGCTTATGATGATCTACGTCAGACAGATACG----- 1591
Qy      501  GluAlaSerAspAlaGluValaLeuAlaSerGluProValaArgAlaTrpPheAlaAspTrp 520
Db      1592  -----GTAGCATCCGCGAGTGG 1609
Qy      521  LeuTyraArgLeuAsnArgGluAlaThrGlyAsnAlaSerArgIleMetTrpVal----- 538
Db      1610  CTC-----TCTATACAACTTGGCGATTACGGG 1636
Qy      539  -----GlyLeuLeuAspThr-----ProProSerIleAspLysGlyGlu 551
Db      1637  AAGAAGTGATGCACTTGATATGACCCCAAGTACCGCACCTAAC----- 1681
Qy      552  ValThrAspLysGlySerIleAsnGlnArgAla-----ValLeuGlnTrpArgSer 566
Db      1682  -----AATTGTTCAAGCCGAGATCGGCTTCCCTGTTTGGCAATGGCGGTCG 1729
Qy      569  AlaLysValaAspAlaLeuTyrrArgGlyGluAspGlnSerMetLeuArgAspGluAlaThr 588
Db      1730  GCGAAAGTTATGCGCTGTATCGTGTGGAAGATCAATCCATCTGCTGACGAGGCCACA 1789
Qy      589  Leu 589
Db      1790  CTG 1792

RESULT 4
AAA39722
ID AAA39722 standard; DNA; 1203 BP.
AC AAA39722;
XX
XX
DT 21-SEP-2000 (first entry)
XX
XX Pseudomonas sp. HR199 fcs-delta DNA.
XX
XX Eugenol; ferulic acid; coniferyl alcohol; coniferyl aldehyde; vanillin;
XX vanillinic acid; flavouring; ferulic acid-CoA-synthetase; fcs; fcs-delta;
XX de.
XX Pseudomonas sp.
XX
XX DE19850242-Al.
XX
XX 04-MAY-2000.
XX
XX 31-OCT-1998; 98DE-01050242.
XX
XX 31-OCT-1998; 98DE-01050242.
XX
XX (HAAR ) HAARMANN & REIMER GMBH.
XX
XX Rabenhorst J, Steinduechel A, Priefert H, Overhage J;
XX
XX WPI; 2000-340642/30.
XX
XX P-PSDB; AA167972.
XX
XX Transformed organisms, used for production of vanillin and other
XX methoxyphenols, have altered catabolism of eugenol or ferulic acid.
XX
XX Claim 8; Fig 21; 80pp; German.
XX
XX This invention describes novel transformed and/or mutated uni- or multi-
XX cellular organisms (A) in which enzymes (I) involved in catabolism of
XX eugenol (E) and/or ferulic acid (FA) are inactivated in such a way that

```

```

QY 372 ProGlyCyseGluValIysLeuValProValAspGlyIlyseuGluGlyArgPheHisGly 391
Db 1158 -----GACTGGCTGCTATTGGCGCGAAGTCCGCGGG 1187
QY 392 ProHisValMetSerGlyTyrTrpArgAlaProGluGlnAsnAlaGlnAlaPheAspGlu 411
Db 1188 CAGGATCTCTGTCATCTAC-----CTTGCTCTCCGCGAAGAAATATCCATCATG----- 1238
QY 412 GluGlyTyrTrpCysSerGlyAspAlaIleLeuValLeuAlaAspProAlaAspPro----- 429
Db 1239 -----GCTGATGCATGCGCGCGCTGCATACGTTGATCCGGCTACC 1280
QY 430 -----GlnIys 431
Db 1281 TGCCCATTCGACCAACAGGAAACATCGCATGAGCAGCAGCATCTCGATGGAAGCC 1340
QY 432 GlyLeuMetPheAspGlyArgIleAlaAspPheIysLeuSerSerGlyValAlaPheVal 451
Db 1341 GGTCTTGTGATCAGATGATCTGAGCAGAAAGACATCAG-----GGGCTCGCCCA 1391
QY 452 SerValGlyProLeuArgThrArgAlaValLeuGluGlyIlySerTyrValLeuAspVal 471
Db 1392 GCCGACGTGTCGCCAGGCTCAAGGCGCGCATGCGCGAG-----GATCTC 1442
QY 472 ValValAlaAlaPheAspArgGlyIlyCysLeuGlyLeuLeuVal-----PhePro 487
Db 1443 GTCTGATACCCATGCGCATGCTCTGCGCAATATGCTGTGAAATAGCGCTTTTCT 1502
QY 488 ArgLeuLeuAspCysArgAlaLeuSerGlyLeuGlyIlyAsnAlaSerAspAlaGluVal 507
Db 1503 GAGTTATCATCTGTGCGCGCGCTG---GCTGTGCGGAGCCGCTATCAGACATACGCTTG 1559
QY 508 LeuAlaSerGluProValAlaGalaTTrpPheAlaAspTrpLeuIysArg-----Leu 524
Db 1560 GCTACCCGTATATTCCTGGAAGAGCTTGCGCGCAATGGGCTACCGCTTCTCTGCTTC 1619
QY 525 AsnArgGluAlaThrArgAlaAsnAlaSerArgIleMetCysAlaGlyLeuLeuAsp-ThrPr 544
Db 1620 TACGGTATCCCGCTCCCGATTCGACAGCGCATCGCTTCTATCGCTTTTGACGAGTTC 1679
QY 544 oProSerIleAspIlyGlyIlyValThrAspIlyGlySerIleAsnGlnArgAlaValLe 564
Db 1680 TTCTGACCGGAGCTCTGGGCTTGAATATGACCAAGC-----GAGCCCTGTTT 1733
QY 564 uGlnTrpArgSerAlaIlyValAlaPalaLeuTyrArgGlyGluAspGlnSerMetLeuAr 584
Db 1734 GCAATGCGCGCTCGCGAAGATTGATGCGTGTATCGTGTGAAGATCAATCCATGCTGCG 1793
QY 584 GAspGluAlaThrLeu 589
Db 1794 TGACGAGGCCACACTG 1809

```

RESULT 3
AAA39721 standard; DNA; 2171 BP.

```

XX AAA39721;
XX
XX 21-SEP-2000 (first entry)
XX
XX Pseudomonas sp. HR199 fcs-omega-Gm DNA.
XX
XX Eugenol; ferulic acid; coniferyl alcohol; coniferyl aldehyde; vanillin;
XX vanillic acid; flavouring; ferulic acid-CoA-synthetase; fcs;
XX fcs-omega-Gm; ds.
XX
XX Pseudomonas sp.
XX
XX DE19850242-A1.
XX
XX 04-MAY-2000.
XX

```

```

PF 31-OCT-1998; 98DE-01050242.
XX
XX 31-OCT-1998; 98DE-01050242.
XX
XX (HAAR) HAARMANN & REIMER GMBH.
XX
XX Radenhorst J, Steinbuechel A, Priefert H, Overhage J;
XX P-PSDB; AA187971.
XX
XX WPI; 2000-340642/30.
XX
XX Transformed organisms, used for production of vanillin and other
XX methoxyphenols, have altered catabolism of eugenol or ferulic acid.
XX
XX Claim 8, Fig 2h; 80pp; German.
XX
XX This invention describes novel transformed and/or mutated uni- or multi-
XX cellular organisms (A) in which enzymes (I) involved in catabolism of
XX eugenol (E) and/or ferulic acid (FA) are inactivated in such a way that
XX the organism accumulates the intermediates coniferyl alcohol (CAI),
XX coniferyl aldehyde (CA2), FA, vanillin (V) and/or vanillic acid (VA).
XX INDEPENDENT CLAIMS are also included for the following: (a) gene
XX structures in which sequences that encode one or more of the enzymes CAI-
XX dehydrogenase, CA2-dehydrogenase, FA-CoA-synthetase, enoyl-CoA hydratase-
XX aldolase, beta-ketothiolase, V-dehydrogenase and VA-demethylase have been
XX altered and/or inactivated; (b) vector containing at least one gene
XX structure of (a); (c) biotechnical production of organic compounds
XX (particularly alcohols, aldehydes or organic acids) using (A); and (d)
XX production method of (A). (A) are used to prepare organic compounds,
XX preferably alcohols, aldehydes and organic acids, especially CAI, CA2,
XX FA, V and/or VA, most particularly V, a well-known natural flavouring.
XX The inexpensive starting material eugenol can be converted to V in a
XX single step. This sequence encodes the Pseudomonas sp. HR199 ferulic acid
XX -CoA-synthetase, fcs-omega-Gm which is described in the method of the
XX invention
XX
XX Sequence 2171 BP; 443 A; 591 C; 616 G; 521 T; 0 U; 0 Other;
XX
XX Alignment Scores:
XX Pred. No. 1,818-59 Length: 2171
XX Score: 766.50 Matches: 227
XX Percent Similarity: 42.75% Conservative: 47
XX Best Local Similarity: 35.41% Mismatches: 158
XX Query Match: 24.87% Indels: 210
XX DB: 3 Gaps: 21
XX
XX US-09-750-986D-30 (1-589) x AAA39721 (1-2171)
QY 1 MetArgSerLeuGluAlaLeuLeuProPheProGlyArgIleLeuGluArgLeuGluHis 20
Db 342 ATTCGTTCTCTGAGGCGCTTCTTCCCTTCCCGGTCGAAATCTTAAAGCTCTCGAGCAT 401
QY 21 TrpAlaIysThrArgProGluGlnThrCysValAlaAlaArgAlaAlaAsnGlyIuTrp 40
Db 402 TGGGCTAAGACCCGTCAGAAACAAACCTGCGTTGTCGACAGGCGGCAATGGGGAATGG 461
QY 41 ArgArgIleSerTyrAlaGluMetPheHisAsnValArgAlaIleAlaGlnSerLeuLeu 60
Db 462 CGTCGTATCAGCTACCGGAAAGTTCACACAGCTCGGCGCATTCGACAGACTTGCTT 521
QY 61 ProTyrGlyLeuSerAlaGluArgProLeuLeuIleValSerGlyAsnAspLeuGluHis 80
Db 522 CTTTACGACTATCGGCAAGAGCTCCGCTGCTTATGCTCTCGAATATACCTCGAATCAT 581
QY 81 LeuGlnLeuAlaPheGlyAlaMetTyrAlaGlyIleProTyrCysProValSerProAla 100
Db 582 CTTCAGCTGGCAATTTGGGCTATGTATGCGGCGCATTCCTATATGCCGCTGCTCTGCT 641
QY 101 TyrSerLeuLeuSerGlnAspLeuAlaIlyValAsnArgHisIleValGlyLeuGluGlnPro 120
Db 642 TATTCACTGCTGCGAAGATTGGCGAAGCTGCGTCAATCGTATGAGTCTTTCGCAACCG 701
QY 121 GlyLeuValPheAlaAlaAspAlaAlaPheGlnArgAlaIleGluThrIleLeuPro 140

```

Db 1870 CTG 1872
RESULT 2
AAA39720
ID AAA39720 standard; DNA; 2188 BP.
XX
AC AAA39720;
XX
DT 21-SEP-2000 (first entry)
XX
XX Pseudomonas sp. HR199 fcs-omega-Km DNA.
DE
XX Eugenol; ferulic acid; coniferyl alcohol; coniferyl aldehyde; vanillin;
KM vanillic acid; flavouring; ferulic acid-CoA-synthetase; fcs;
XX fcs-omega-Km; ds.
XX Pseudomonas sp.
OS
XX DE19850242-A1.
XX
XX 04-MAY-2000.
XX
XX 31-OCT-1998; 98DE-01050242.
XX
XX 31-OCT-1998; 98DE-01050242.
XX
XX 31-OCT-1998; 98DE-01050242.
XX
XX (HAAR) HAARMANN & REIMER GMBH.
XX
XX Rabenhorst J, Steinduechel A, Priefert H, Overhage J;
PI WPI; 2000-340642/30.
XX P-PSDB; AAY87970.
DR
XX
XX Transformed organisms, used for production of vanillin and other
PT methoxyphenols, have altered catabolism of eugenol or ferulic acid.
XX
XX
XX Claim 8; Fig.2g; 80pp; German.
XX
XX This invention describes novel transformed and/or mutated uni- or multi-
CC cellular organisms (A) in which enzymes (I) involved in catabolism of
CC eugenol (E) and/or ferulic acid (FA) are inactivated in such a way that
CC the organism accumulates the intermediates coniferyl alcohol (CA1),
CC coniferyl aldehyde (CA2), FA, vanillin (V) and/or vanillic acid (VA).
CC INDEPENDENT CLAIMS are also included for the following: (a) gene
CC structures in which sequences that encode one or more of the enzymes CA1-
CC dehydrogenase, CA2-dehydrogenase, FA-CoA-synthetase, enoyl-CoA hydratase-
CC aldolase, beta-ketochiolase, V-dehydrogenase and VA-demethylase have been
CC altered and/or inactivated; (b) vector containing at least one gene
CC structure of (a); (c) biotechnical production of organic compounds
CC (particularly alcohols, aldehydes or organic acids) using (A); and (d)
CC production method of (A). (A) are used to prepare organic compounds,
CC preferably alcohols, aldehydes and organic acids, especially CA1, CA2,
CC FA, V and/or VA, most particularly V, a well-known natural flavouring.
CC The inexpensive starting material eugenol can be converted to V in a
CC single step. This sequence encodes the Pseudomonas sp. HR199 ferulic acid
CC CoA-synthetase, fcs-omega-Km which is described in the method of the
CC invention
XX
XX
SQ Sequence 2188 BP; 408 A; 607 C; 670 G; 503 T; 0 U; 0 Other;
XX
Alignment Scores:
Pred. No.: 2,04e-62 Length: 2188
Score: 799.00 Matches: 234
Percent Similarity: 45.53% Conservative: 51
Best Local Similarity: 37.38% Mismatches: 168
Query Match: 25.92% Indels: 173
Gaps: 20
US-09-750-986D-30 (1-589) x AAA39720 (1-2188)
QY 1 MetArgSerLeuGluAlaLeuLeuProPheProGlyValGlyIleLeuGluArgLeuGluHis 20
|||||

Db 342 ATGCGTTCTCTCGAGCGCTTCTCCCTTCGCCGGTCCGAATCTTTCAGCTTCGACCAT 401
QY 21 TPAlaIySerThraArgProGluGlnThrCysValAlaAlaArgAlaAlaAngIyGluTrp 40
Db 402 TGGGCTTAAGACCCGCTCCAGAACAAACCTGGTTGCTCCAGGGCGGCAAAATGGGAATGG 461
QY 41 ArgArgIleSerTyraAlaGluMetPheHisAsnValArgAlaIleAlaGlnSerLeuLeu 60
Db 462 CGTCGATACAGCTACGCGGAATGTTCCACAAAGCTCCGCCATTCGCACAGACTTGCTT 521
QY 61 ProTyrgIlyLeuSerAlaGluArgProLeuLeuIleValSerGlyAsnAspLeuGluHis 80
Db 522 CCTTACGGACTATCGCAGAGCGCTCGCTTATGCTCTGGAATGACCTGGAACAT 561
QY 81 LeuGlnLeuAlaPheGlyAlaMetTyraAlaGlyIleProTyrcysProValSerProAla 100
Db 582 CTTACAGTGGCATTTGGGCGCTATGATCGGGCAATTCCTTATGCCCGGTCTTCCTGCT 641
QY 101 TyrsSerLeuLeuSerGlnAspLeuAlaIyLeuArgHisIleValGlyLeuLeuGlnPro 120
Db 642 TATTCACGTGCTCGCAAGATTGGCGAGCTCGTCACATCGTAGGCTTTCGCAACCG 701
QY 121 GlyLeuValPheAlaAlaAspAlaAlaProPheGlnArgAlaIleGluThrIleLeuPro 140
Db 702 GGACTGGCTTCTTCTGCTCCGATGCGACACCTTTCAGGG- 740
QY 141 AspAspValProAlaIlePheThrArgGlyGluLeuAlaGlyArgArgThrValSerPhe 160
Db 740 - 740
QY 161 AspSerLeuLeuGlnGlnProGlyGlyIleGluAlaAspAsnAlaPheAlaAlaThrGly 180
Db 740 - 740
QY 181 ProAspThrIleAlaIySphenLeuPheThrSerGlySerThrIySleuProIySAlaVal 200
Db 741 - ACAGCAGCGAACCAGAAATTCGACGCGGGCG 773
QY 201 ProThrThrGlnArgMetLeuCysAlaAsnGlnGlnMetLeuLeuGlnThrPheProVal 220
Db 774 CCTCTCGTAG- 785
QY 221 PheGlyGluGlnProProValLeuValAspTrp- -LeuPro- 233
Db 786 GTTGGGAG- -CCCTGCAAAATTAACCTGATGGCTTCTTCCGCCCAAGATCTGA 839
QY 234 TPAsnHisThrPheGlyGlySerHisAsnIleGlyIleValLeuTyraAngIyGlyThr 253
Db 840 TGGCGCAGGGGATCAAGATCTGATCAAGACAGGA- -TGAGATCG 884
QY 254 TyrTyrlLeu- -AspAspGlyIySProThrAlaGlnIyPheAlaGluThrLeuArg 271
Db 885 TTTCGATGATTTAAACAAGATGA- 908
QY 272 AsnLeuSerGlnIleSerProThrAlaTyrlLeuThrValProIySgIyTrpGluGluLeu 291
Db 909 - -TTGACGCGAGCTTCTCGGCGCTTGGGTGAGAGAGCTATTTCGCTATGACTGGGCA 965
QY 292 ValGlyAlaLeuGlnArgAspSerThrLeuArgGluArgPhePheAlaAngMetIySleu 311
Db 966 CAACAGACATCGGCTGCTGTGATGCCCGCGTTCGCTGTCAGCGAGGGCGCCG 1025
QY 312 PhePhePheAlaAlaIaGlyLeuSerGlnGlyIleTyraAspArgLeuAspArgValAla 331
Db 1026 GTTCTTTTGTCAAGACCGACCTGTCGGGCGCTTAATGAACTGCAGACGAGCGAGCG 1085
QY 332 GluGlnHisCysGlyGluArgIleArgMetMetAlaGlyLeuGlyMetThrGluThrAla 351
Db 1086 - -GGGCTATCGTGGCTGCGCCAGCGAGCGGCTT- 1115
QY 352 ProSerCysThrPheThrThrGlyIProLeuSerMetAlaGlyTyrlIleGlyLeuProAla 371
Db 1116 - -CCTTGGCGACGCTGCTGCTGCACTGTTCACAGAACGCGGAGG- 1157

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 7, 2005, 23:50:05 ; Search time 669 Seconds
(without alignments)
4621.688 Million cell updates/sec

Title: US-09-750-986D-30
Perfect score: 3082
Sequence: 1 MRSLEALPPGRILRLERH.....KVDALYRGEDQSMRLDEATL 589

Scoring table: BLASTSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 413486 seqs, 2624710521 residues
Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xid
-Q=/cgrn2.1/USPTO.spool_P/US09750966/runat_03022005_071743_18555/app_query.fasta_1.775
-DB=N.Geneseq_23Sep04 -QFMT=fastap -SUFFIX=trng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNIT=bits -START=1 -END=-1 -MATRIX=blomsum62 -TRANS=human40.cdi
-LIST=45 -DOCMATCH=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=plto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09750966 @CGN 1.1 708 @runat 03022005 071743 18555 -NCPU=6 -ICPU=3
-NO MAMP -IAREBOJERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEF_TIMEOUT=120 -WARN_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: Geneseqn1980s:*
2: Geneseqn1980s:*
3: Geneseqn2000s:*
4: Geneseqn2000s:*
5: Geneseqn2000s:*
6: Geneseqn2000s:*
7: Geneseqn2000s:*
8: Geneseqn2000s:*
9: Geneseqn2000s:*
10: Geneseqn2000s:*
11: Geneseqn2000s:*
12: Geneseqn2000s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1652.5	53.6	1902	9 ADA31106	ADA31106 DNA encod
2	799	25.9	2188	3 AAA39720	AAA39720 Pseudomon
3	766.5	24.9	2171	3 AAA39721	AAA39721 Pseudomon
4	698.5	22.7	1203	3 AAA39722	AAA39722 Pseudomon
5	314.5	10.2	349980	6 ABO81844	ABQ81844 Bifidobac
6	309	10.0	38494	10 ADB74275	ADB74275 Mycobacte

7	3021.5	9.8	35961	10 ADB74389	ADB74389 Mycobacte
8	301.5	9.8	37286	4 AAS59522	AAS59522 Propionib
9	301.5	9.8	37286	4 ACF64451	ACF64451 Propionib
10	290	9.4	1803	10 ACF71666	ACF71666 Photorhab
11	290	9.4	110000	10 ACF67367	ACF67367 Photorhab
12	284	9.2	2103	11 ADL72353	ADL72353 Modified
13	283	9.2	2103	12 ADN72468	ADN72468 Thale cte
14	283	9.2	110000	2 AAT42063	AAT42063 Haemophil
15	279	9.1	2103	7 ADL72371	ADL72371 Plant acy
16	279	9.1	2103	11 ADL72249	ADL72249 Arabidops
17	278.5	9.0	349980	6 ABO81849	ABQ81849 Bifidobac
18	278	9.0	2106	6 ADL23370	ADL23370 Plant acy
19	277	9.0	2097	6 ABZ14464	ABZ14464 Arabidops
20	277	9.0	2097	6 ADG87806	ADG87806 A. thalia
21	277	9.0	2097	8 ADA67967	ADA67967 Arabidops
22	275	8.9	2257	11 ADL72248	ADL72248 Arabidops
23	268.5	8.7	349980	6 ABO81846	ABQ81846 Bifidobac
24	268	8.7	2262	3 AAC42047	AAC42047 Arabidops
25	267	8.7	2154	8 ADA70529	ADA70529 Rice gene
26	266.5	8.6	2094	8 ADA69386	ADA69386 Rice gene
27	266	8.6	1845	10 ADF02609	ADF02609 Bacterial
28	265.5	8.6	85692	12 ADI39159	ADI39159 Streptomy
29	264.5	8.6	1639	2 AAX32852	AAX32852 Mutant lu
30	263.5	8.5	1639	4 AAS00899	AAS00899 Beetle cd
31	263.5	8.5	1639	4 AAS00894	AAS00894 Beetle cd
32	263.5	8.5	1639	6 ABO81820	ABQ81820 Photuris
33	263.5	8.5	1639	12 ADM32125	ADM32125 Mutant lu
34	263	8.5	3264	10 ABV75591	ABV75591 Novel hum
35	263	8.5	6162	12 ADN04241	ADN04241 Antipori
36	262.5	8.5	1639	2 AAX32853	AAX32853 Mutant lu
37	262.5	8.5	1639	4 AAS00895	AAS00895 Beetle cd
38	262.5	8.5	1639	6 ABO81818	ABQ81818 Photuris
39	262.5	8.5	1639	12 ADM32123	ADM32123 Mutant lu
40	262.5	8.5	12951	8 ACA43808	ACA43808 Prokaryot
41	259	8.4	110000	4 AAI99682	AAI99682 (25 o
42	259	8.4	110000	4 AAI99683	AAI99683 (25 o
43	258	8.4	1639	4 AAS00898	AAS00898 Beetle cd
44	258	8.4	1639	6 ABO81819	ABQ81819 Photuris
45	258	8.4	1639	12 ADM32124	ADM32124 Mutant lu

ALIGNMENTS

RESULT 1	
ID ADA31106	ADA31106 standard; DNA; 1902 BP.
AC ADA31106;	
XX	
DT 20-NOV-2003	(first entry)
XX	
DE DNA encoding Acinetobacter baumannii protein #2393.	
XX	
KW de; gene; Acinetobacter baumannii; bacterial disease; antibacterial;	
XX vaccine; plant biocontrol agent.	
OS Acinetobacter baumannii.	
XX	
PN US6562958-BI.	
XX	
PD 13-MAY-2003.	
XX	
PF 04-JUN-1999;	99US-00328352.
XX	
PR 09-JUN-1998;	98US-0088701P.
XX	
PA (GENO-) GENOME THERAPEUTICS CORP.	
XX	
PI Breton G, Bush D;	
XX	
DR WPI; 2003-576092/54.	
XX	
DR P-PDB; ADA35232.	
XX	

This Page Blank (uspto)

```
Db 170 FT-----LIYTSGETGEPKGMIDYONIAA--QLYHDOQLTTEODISISFLPLS 218
Qy 236 HTEGGSNIGIVLYNGSTVYLDGKPTAQGFAETLRNLSEISPTAYLTVPKMEELVGA- 294
Db 219 HIFERAMSYVMHTGALNVLTDNPFREA-----MLDVRPTVMGSVRLYEKIHSAI 271
Qy 295 LERDSTLRERFFAPAMKLFPPAAGLSQGIWDR-----LDRVA---EQ 333
Db 272 LEKVS--RAPFWR--IIFQWAIKRGERRRMSRQINDKSCPFSLCGEKLADRLVINKLRQ 327
Qy 334 HCGERTRMMA-----GLGWTETAPSGCTFTTGPLSMAGYIGLP 370
Db 328 LILGGRIRFLVAGARLEDSIMRFFLSVGLNIKYGYGTETCATVSCWEERKNYILGSGITP 387
Qy 371 APGCEVKLYPVVDGKLEGRFHPHVMGSGWRAPESONAQAFDEGYCSGDAIKLADPADPQ 430
Db 388 LPGVDVR-ISTAGEIQVR--GPIVMKGYNRPEDTVNHFTEDEGWLRTGDAGL---DET 440
Qy 431 KGLMPDGRITAEDFKLSGCVFVSVPLEKTRAVLEGGSYLDVVVAAPDRECLGLIVPPRL 490
Db 441 GNLFIATERLKDLMKTSYGKYIA--POLIEGTLGQDRFIEHIAVIADARQFVSALIYP--- 495
Qy 491 DCRALSGIGKE-----ASDAEVLASEPVRAMPADWLKRLNREATGNASRIMWTGLDTPP 545
Db 496 SYEALEDYAKSINLHYQNRLLELNHNOVIGLFEORLKELOKN-IAKFHQVKKRFTLLPESF 554
Qy 546 SIDKGEVTDKGSINORAVLQWRSKVDALYR 576
Db 555 SMEKGELTPTLKLRRKVIISQRYQSEIESMYR 585
```

Search completed: February 7, 2005, 23:59:47
job time : 109 secs

```

Db 302 ----SDVYISYLPPLAHYERANOIMGVYG---VAVGFYQGDVFKLMD----- 342
Qy 271 RNLSSEIPTAYLTPVPGWELVGLERDSTLRERFPFARMKLPFAAAGLSOGI----- 323
Db 343 -DEAVLRPTIFCSVPLRYNRITDGI--TSAVKSSGVKKRLFEIAYNSKKQALINRTPS 399
Qy 324 --WDRU--DRVAEOHGERIRMA-----GLGNTETAPSGT 355
Db 400 AFMDKLVFNKIKK-LGGRVRFMGSGASPLSPVMDFLRICFGCSVREGYMET--SCV 456
Qy 356 FTT---GPLSMAGYIGLPAPGCEVKLVPD-----GKLEGFPHGVMSGW 399
Db 457 ISAMDDG-NLSGHVSPNPACCEVKLVDPENMYTSDQPYPRGEICVR--GPILFKGY 513
Qy 400 RAPEQNAQAFDEGYGCSGDAIKLADPADPQKGLMPD-----RIADFKLSSGVFS 452
Db 514 KQEEQREILDDGGMHTGTDI-----GLMLPGRLKTIIDRKNIIFKLAQGEYIA 562
Qy 453 VGFLR-----TRAVLEGGSY---VLDVVVAAPD-----RECLGLLV-PP 487
Db 563 PEKIEVNTYKCRFVSCGFHGDSPNSLVAIVSDPEVKMDMAASGIKYEHGQLCNDP 622
Qy 488 RLID--CRALSGLGEA-----SDAEVLASEP-----VRAMPADW 520
Db 623 RVAKTVLAEMDDIGREAOGRGFEFAKAVTLVPEPFTLENGLLPTPFKIKRPAKAYFAEA 682
Qy 521 LKRLNE 527
Db 683 ISKRYAE 689

```

RESULT 14

```

US-10-410-031-130
; Sequence 130, Application US/10410031
; Publication No. US20040010817A1
; GENERAL INFORMATION:
; APPLICANT: Shockley, Jay M.
; APPLICANT: Schmutz, Judy
; APPLICANT: Browne, John A.
; TITLE OF INVENTION: Plant Acyl-CoA Synthetases
; FILE REFERENCE: DOM-07654
; CURRENT APPLICATION NUMBER: US/10/410,031
; CURRENT FILING DATE: 2003-04-09
; NUMBER OF SEQ ID NOS: 191
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 130
; LENGTH: 700
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-410-031-130

```

```

Query Match 9.2%; Score 283; DB 15; Length 700;
Best Local Similarity 22.5%; Pred. No. 8.8e-18;
Matches 150; Conservative 97; Mismatches 214; Indels 206; Gaps 35;

```

```

Db 302 ----SDVYISYLPPLAHYERANOIMGVYG---VAVGFYQGDVFKLMD----- 342
Qy 271 RNLSSEIPTAYLTPVPGWELVGLERDSTLRERFPFARMKLPFAAAGLSOGI----- 323
Db 343 -DEAVLRPTIFCSVPLRYNRITDGI--TSAVKSSGVKKRLFEIAYNSKKQALINRTPS 399
Qy 324 --WDRU--DRVAEOHGERIRMA-----GLGNTETAPSGT 355
Db 400 AFMDKLVFNKIKK-LGGRVRFMGSGASPLSPVMDFLRICFGCSVREGYMET--SCV 456
Qy 356 FTT---GPLSMAGYIGLPAPGCEVKLVPD-----GKLEGFPHGVMSGW 399
Db 457 ISAMDDG-NLSGHVSPNPACCEVKLVDPENMYTSDQPYPRGEICVR--GPILFKGY 513
Qy 400 RAPEQNAQAFDEGYGCSGDAIKLADPADPQKGLMPD-----RIADFKLSSGVFS 452
Db 514 KQEEQREILDDGGMHTGTDI-----GLMLPGRLKTIIDRKNIIFKLAQGEYIA 562
Qy 453 VGFLR-----TRAVLEGGSY---VLDVVVAAPD-----RECLGLLV-PP 487
Db 563 PEKIEVNTYKCRFVSCGFHGDSPNSLVAIVSDPEVKMDMAASGIKYEHGQLCNDP 622
Qy 488 RLID--CRALSGLGEA-----SDAEVLASEP-----VRAMPADW 520
Db 623 RVAKTVLAEMDDIGREAOGRGFEFAKAVTLVPEPFTLENGLLPTPFKIKRPAKAYFAEA 682
Qy 521 LKRLNE 527
Db 683 ISKRYAE 689

```

RESULT 15

```

US-10-369-493-311
; Sequence 311, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 311
; LENGTH: 585
; TYPE: PRT
; ORGANISM: Xenorhabdus nematophilus
US-10-369-493-311

```

```

Query Match 9.1%; Score 281; DB 14; Length 585;
Best Local Similarity 22.2%; Pred. No. 1.1e-17;
Matches 140; Conservative 105; Mismatches 268; Indels 118; Gaps 21;

```

```

Qy 9 PEPGRILEREHNAKTRPEQTCVAARAAN-----GEMRIRISYAEFHNVAIAQSLLPYGL 64
Db 81 PRIGTLHDFVHAVENTAENKYLSTRVSDGTIGETISMMTYGEASERQALISGLLFHGV 140
Qy 65 S-----AERPLIVSGNDLEHLQALFAGMAYAGIPYCPVSPAVSLISQDLAK--LRH 113
Db 141 NQDDCVGLYFIRNPEMV-----VDHCAAYSFSV-----PLYDTGLPDVAKFVYVNH 188
Qy 114 -----IVGLIQGLVPAADAAPFORALETI--LPDDVPAIFTRGELAGRTVSPDSL 163
Db 189 ANLOAIFCVQPTNILLISFLAE--IPSIRLIWVGADENHPSL-PRG--TGVIVTSYOKL 244
Qy 164 LEOPGIEANNAFAANGSDTIATKLFSSGSTKLKPAVPTQRMICAN---QOMLQTFPV 220
Db 245 LSG--GRSSIAHPSPPKPEBIAITICYSGTGTPKGVLLHGNLIANVASSVEAEFP- 301
Qy 221 FGESEPLVLMPLMNH-----TFGGSNIGIYLVNGGTYVLDGKPTAQGFATL 270

```

```

Qy 18 LEHWATREBQTCVAAARAANGEMRIRISYAEFHNVAIAQSLLPYGLSAERPLIVSGN- 76
Db 1 LQQQVSKYEKELIAFRQMSAS--EQLKMSQWDVDSKTSIARSLSLSEVGIQEKIGIPADS 59
Qy 77 -----DLEHLQALFAGMAYAGIPYCPVSPAVSLISQDLA-----KLRIHVLGQ 121
Db 60 MAWSLSDLLILQAR-----AVTVPLVATSSRDOAFILNDAGVAILFVGGQEQYD 109
Qy 122 LVFAADAAPFORALETILPDDV-----PAIFTRGELAGRTVSPDSLQOPGIEADNA 175
Db 110 IALILALCPQLNHLVVLDESVDLKECPQALYISATIADRPYQAELELSRAGLDIHL 169
Qy 176 FAATGPTIATKLFSSGSTKLKPAVPTQRMICANQOMLQTFPVFGESEPLVLMPLMNH 235

```

Qy	298	DSTLRERFARKL-----FFFA----	LAG-----LSGG--	IMDLDRPAEHCGERIR	340
Db	295	DMQKNQGFFFNYQLTKANKIIFSSKWEALGAGNVITVSGAALQPRLAR--	FMAAQIP	351	
Qy	341	MMAGLGMTETAPSCTF-----	TTGPTLSMAGYIGLPABGCWVKLVVPGKLEGRF	389	
Db	352	VMEGGGLTETSIVLIANKRPNDPETRISTYGPL-----	LPGEVVKIAE-DGEILTR-	400	
Qy	390	HGPHVMSGYWRAPBEONAAQFDEEGYCCSDAIKLADPADPOKGLMFDGRIABDFKLSGV	449		
Db	401	-GPHIMKGYKKRPDLTAEALDSEGWHGTIDIQGFVE----	NKTLKLTTRDKKEHFKNSGK	455	
Qy	450	FVSVGPELTRAVLEGGSYLVLDVVAAAPDRECTGLVFPRLDGRA--	LSGLGKASDAE	506	
Db	456	YIA--PQLVENKFKESRYIEQLMVTIGEGEKKFSPSAIVVSPFADVKSCEIKGI-	PLYTTEE	512	
Qy	507	VLASPPVAMPAPDWLKLRLNREATGNASRLTMVGLDTPPSIDKGEVTTDKGSIINQRAVLQW	566		
Db	513	MVCNAQVIEKQKEDENM-ENFAQYERVKETILMPRPETIEGEMTPTLKLKRIKTIOMN	571		
Qy	567	RSAKVDALY	575		
Db	572	CQGQLEKMY	580		

```

RESULT 12
US-09-906-419-17
; Sequence 17, Application US/09906419
; Publication No. US20030037357A1
; GENERAL INFORMATION:
; APPLICANT: Schockey, Jay
; APPLICANT: Schnurr, Judy
; APPLICANT: Browne, John
; TITLE OF INVENTION: Plant Acyl-CoA Synthetases
; FILE REFERENCE: DOW-04679
; CURRENT APPLICATION NUMBER: US/09/906,419
; CURRENT FILING DATE: 2001-07-16
; PRIOR APPLICATION NUMBER: 60/220,474
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 17
; LENGTH: 700
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; US-09-906-419-17

```

Query Match	9.2%;	Score 283;	DB 10;	Length 700;
Best Local Similarity	22.5%;	Pred. No. 8.8e-18;		
Matches 150;	Conservative 97;	Mismatches 214;	Indels 206;	Gaps 35;

```

QY      9  FPPGGILERLEMAWTRPQGTQVAAARAAN-----GEWRRISYEMEFNNVAIAQSLIPYGL 64
Db      81  PEIGLHNPVHAVENTYKENKYLGRVNSDGTIGEISMTYGEHBAERQAGISGLIFHG 140
QY      65  S-----AERPLLIYSGNDLEHQLAFGAMYAGIPLYCPVSPAYSLISODLAK--LRR 113
Db      141  NQGDCEVLGYFINRPEMLV-----VDHCAKAYSFVSV-----PLVDLTGPDVAKFVANH 188
QY      114  -----IVGLDQGLVFPADAAFPQRAIETT--LPDDVPALFTRGELAGRRTVSFDSL 163
Db      189  ANLQAFICVPQTLNLTLSFLAE--ISIRIIVVVGADENHPLSL-PRG--TGVTIVSYOKX 244
QY      164  LEOPGIEADNAAFAATGPDITAKPLFTPSGSKLPRKAVPTQOMLCAN-----QQMLLTQFPV 220
Db      245  LSQ--GRSSLHFSPKPEDIAITTCYTSQITGTTPKVVLTGHLINNAVAGSSVEAFEPF 301
QY      221  FGEBPPVLVDMLPMNH-----TFGSGHNIGIVLYNGGTYVLDDEKPTAQGEAETL 270
Db      302  ----SDVYISYLPFLHIYERANQINGVGG--VANGFQGVDFKLM----- 342
QY      271  RNLSIEISTAYLTVYKGMELVGALEKSDTLERRFPARKLPPFAAAGLSQI----- 323

```

[illegible]

```

RESULT 13
US-10-119-136-130
; Sequence 130: Application US/101119136
; Publication No. US20030097676A1
; GENERAL INFORMATION:
; APPLICANT: Shockeey, Jay
; APPLICANT: Schmutz, Judy
; APPLICANT: Browne, John
; TITLE OF INVENTION: Plant Acyl-CoA Synthetases
; FILE REFERENCE: DOW-04695
; CURRENT APPLICATION NUMBER: US/10/119,136
; PRIOR FILING DATE: 2002-09-04
; PRIOR APPLICATION NUMBER: 09/906,419
; PRIOR FILING DATE: 2001-07-16
; PRIOR APPLICATION NUMBER: 60/220,474
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 132
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 130
; LENGTH: 700
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-119-136-130

```

Query Match	9.2%	Score 283;	DB 14;	Length 700;
Best Local Similarity	22.5%;	Pred. No. 8.8e-18;		
Matches 150;	Conservative 97;	Mismatches 214;	Indels 206;	Gaps 35;

```

0Y 9 PEPGGILERLEHMAATREQCVCVAAARAAN-----GEMRRTSYAEMFNFNAIAKOSLLPYGJ 64
Db 81 PEIGLHNFHVAHVTYAKNTKLTGRVNSDGTIGESMWTYEAASERQAIISGGLFHCV 140
0Y 65 S-----ABRPILIYSGNDLEHLQIAFCAMYAGIPIYCPVSPAYSLLSODIAK--LKH 113
Db 141 NOGDVCVGLYFINRPFWLV-----VDHACAAYSVSU-----PLVDLTGPDAVKRKVVNH 188
0Y 114 -----IVGILLOPGLVPAADAAFPORAIEFI--LPDDVPAIFTRGELAGRRIVSPDSI 163
Db 189 ANLOAIFCVQPTNLNILLSFLAE--IPSIRILVVGGADEHLFSL--PRG--TGWTIVSYOKL 244
0Y 164 LEQGGGIEADNAPATGPDITAKFLPFTSGSTLCPAAPTQOMLCAN---QQMLLOTFPV 220
Db 245 LSQ--GRSLHPFSPPKPEDIATICTYTGTTGTTPKGVULTHGNLIANVAGSSVEAEFFP- 301
0Y 221 FGEBPVLVDMVPMNH-----TFGSHNIGIVLVNGSTYYLIDGKPTAOGFAETL 270

```

RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
 RA Ouriello B., Goldman B.S., Cao Y., Akenazi M., Halling C., Mullin L.,
 RA Hounmel K., Gordon J., Vaudin M., Tatchouk O., Bep A., Liu F.,
 RA Moliam C., Allinger M., Doughty D., Scott C., Lappas B.,
 RA Flanagan C., Crowell C., Gursen J., Lomo C., Sear C., Strub G.,
 RA Cielo C., Slater S.,
 RT "Genome sequence of the plant pathogen and biotechnology agent
 RT Agrobacterium tumefaciens C58."
 RL Science 294:2323-2328(2001).
 CC -1- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
 CC family.
 CC EMBL: AEO08067; AAK87208.1; -
 DR InterPro: IPR000873; AMP-bind.
 DR Pfam: PF00501; AMP-binding.1.
 DR PRINTS: PR00154; AMPBINDING.
 DR PROSITE: PS00455; AMP_BINDING.1.
 DR SQUENCE 591 AA; 65535 MW; 9B371D0064CBE856 CRC64;
 SQ
 Query Match 42.0%; Score 1294; DB 2; Length 591;
 Best Local Similarity 46.5%; Pred. No. 7.8e-86;
 Matches 266; Conservative 84; Mismatches 220; Indels 2; Gaps 2;
 QY 5 EALLPFGRIERLEHNAKTRPEQTCVAAAPANGEMRIRISYAEMFNHVRVIAQSILPYGL 64
 DB 6 DPLGYPQKLNRLHMCRSAPERTWADRGREPMRVSVAALDKIRRIQGLDHL 65
 QY 65 SARRPLIVSGNDLEHQLAFGMYAGIPYCPVSPAYSLISODLAKRHHVGLQPLVF 124
 DB 66 SVSRPLIVSENSIEHALMVAQHVIAASAIAPAYA-TSADTKLADIRGQITPEMVF 124
 QY 125 AAAAPFORAIETILPDDVAIFTRGELAGR-RTVSFDSLLEQPGGIEADNAFAATGPD 183
 DB 125 AEDATPRRALGVEFDGTLVGLRLNPEDRSNTFHETLLETPEPTAVIDRAFAVGPDT 184
 QY 184 IAKFLFTSGSTKLPAKAVPTTORMLCANQMLLOTFPVGEPEPVLDMLPMNHTFGSSHN 243
 DB 185 VAKFLFTSGTTPSKAVIQTORMLCNSQEMIAACYGFREPEPVVDMAPMNHTAANKV 244
 QY 244 IGVLYNGGTYIYDDGKPTAQGFATLRNLSISPTAYLYVPKWEELVGALEBDSTLRE 303
 DB 245 FNLVLYNGGTYIYDRGKPSPAQIQGTLDNRDISPTVFNVPACHHEMLVQAMKDEALCR 304
 QY 304 RFPARMKLFFPFAAGLSQGIWDRLDRAVEQHCGERIRMAAGLGMTETAPSCFTTGPLSM 363
 DB 305 SFPRDLKMLVYAGMAGHTWDALTELSMATVGAIVMGAGLSTERTAPPSLCTEPQDK 364
 QY 364 AGYIGLPAPCEVYLVPVDKLEGRFHGPHVMGMYRABEQNAQAPDEEGYCSGDAIKL 423
 DB 365 PGNIGIPAGQVTMVLVPDGRYELRLKGNPITPGYMNGLTAAAPDEEGYRIGDTVKF 424
 QY 424 ADPADPQKLMFPGRIAEDEKLSGCVSVSGPLFTRAVLEGGSYVLDVVVAAPDEECLGL 483
 DB 425 AVADDPARGFYFDGRMAENFKLOTGTWVAAGPLAQLVNMFAGLIRDAVITGGERALGA 484
 QY 484 LVPEPRLLDCRALSGLKEASDAEVLASEPVRAMPADMLKRLNREATGNASRIMVGLD 543
 DB 485 LVVPEPRIALRELVRGSHLSDAETIRHPSVRAQIVAKLSAQKQASGSASRVMRIIWMED 544
 QY 544 PSIDKGEVTDKGSINQRAVLQWMSAKVDALY 575
 DB 545 ALRFEKGEVTDKGSINQRAVLHRLKELVESLY 576
 RESULT 15
 ID Q8UF11 PRELIMINARY; PRT; 611 AA.
 AC Q8UF11;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Long-chain-fatty-acid-CoA-ligase.
 GN OrderedLocusNames=Atu1416;
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
 OC NCBI_TaxID=176299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Dupont;
 RX MEDLINE=21608550; Pubmed=11743193;
 RA Wood D.W., Setubal J.C., Kaul R., Monke D.E., Kiteajima J.P.,
 RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,
 RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D., Grant C.,
 RA Chapman P., Clendenning J., Deachere G., Gillet W., Grant C.,
 RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmeri A.,
 RA Raymond C., Rouse G., Saenphimachak C., Wu Z., Romero P., Gordon D.,
 RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
 RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
 RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
 RA Nester E.W.,
 RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
 RT C58."
 RL Science 294:2317-2323(2001).
 CC -1- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
 CC family.
 CC EMBL: AEO09102; AAL42422.1; -
 DR PIR: AH2750; AH2750.
 DR PIR: G97531; G97531.
 DR GO: GO:0016874; F.ligase activity; IEA.
 DR GO: GO:0008152; P.metalloisim; IEA.
 DR InterPro: IPR000873; AMP-bind.
 DR Pfam: PF00501; AMP-binding.1.
 DR PRINTS: PR00154; AMPBINDING.
 DR PROSITE: PS00455; AMP_BINDING.1.
 KM Complete proteome; Ligase.
 SQ SEQUENCE 611 AA; 67860 MW; D86FF0A396B05860 CRC64;
 Query Match 42.0%; Score 1294; DB 2; Length 611;
 Best Local Similarity 46.5%; Pred. No. 8.1e-86;
 Matches 266; Conservative 84; Mismatches 220; Indels 2; Gaps 2;
 QY 5 EALLPFGRIERLEHNAKTRPEQTCVAAAPANGEMRIRISYAEMFNHVRVIAQSILPYGL 64
 DB 26 DPLGYPQKLNRLHMCRSAPERTWADRGREPMRVSVAALDKIRRIQGLDHL 85
 QY 65 SARRPLIVSGNDLEHQLAFGMYAGIPYCPVSPAYSLISODLAKRHHVGLQPLVF 124
 DB 86 SVSRPLIVSENSIEHALMVAQHVIAASAIATPAYA-TSADTKLADIRGQITPEMVF 144
 QY 125 AAAAPFORAIETILPDDVAIFTRGELAGR-RTVSFDSLLEQPGGIEADNAFAATGPD 183
 DB 145 AEDATPRRALGVEFDGTLVGLRLNPEDRSNTFHETLLETPEPTAVIDRAFAVGPDT 204
 QY 184 IAKFLFTSGSTKLPAKAVPTTORMLCANQMLLOTFPVGEPEPVLDMLPMNHTFGSSHN 243
 DB 205 VAKFLFTSGTTPSKAVIQTORMLCNSQEMIAACYGFREPEPVVDMAPMNHTAANKV 264
 QY 244 IGVLYNGGTYIYDDGKPTAQGFATLRNLSISPTAYLYVPKWEELVGALEBDSTLRE 303
 DB 265 FNLVLYNGGTYIYDRGKPSPAQIQGTLDNRDISPTVFNVPACHHEMLVQAMKDEALCR 324
 QY 304 RFPARMKLFFPFAAGLSQGIWDRLDRAVEQHCGERIRMAAGLGMTETAPSCFTTGPLSM 363
 DB 325 SFPRDLKMLVYAGMAGHTWDALTELSMATVGAIVMGAGLSTERTAPPSLCTEPQDK 384
 QY 364 AGYIGLPAPCEVYLVPVDKLEGRFHGPHVMGMYRABEQNAQAPDEEGYCSGDAIKL 423
 DB 385 PGNIGIPAGQVTMVLVPDGRYELRLKGNPITPGYMNGLTAAAPDEEGYRIGDTVKF 444
 QY 424 ADPADPQKLMFPGRIAEDEKLSGCVSVSGPLFTRAVLEGGSYVLDVVVAAPDEECLGL 483
 DB 445 AVADDPARGFYFDGRMAENFKLOTGTWVAAGPLAQLVNMFAGLIRDAVITGGERALGA 504
 QY 484 LVPEPRLLDCRALSGLKEASDAEVLASEPVRAMPADMLKRLNREATGNASRIMVGLD 543
 DB 505 LVVPEPRIALRELVRGSHLSDAETIRHPSVRAQIVAKLSAQKQASGSASRVMRIIWMED 564

Qy 544 PSIDKGEVTDKGSINQRAVLQWRSACVDALY 575
:|||||:|||||:|:|
Db 565 ALRFEKGEVTDKGSINQRAVLHLHRELVESTLY 596

Search completed: February 7, 2005, 23:47:44
Job time : 147 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 7, 2005, 23:51:01 | Search time 5877 Seconds
(without alignments)
4739.430 Million cell updates/sec

Title: US-09-750-986D-30

Perfect score: 3082
Sequence: 1 MRSLEALPFGRIILERLEH.....KYDALVRGDSQMLRDEATL 589

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4526729 segs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_plus_p2n.model -DEV=xip
-O=/cgn2.1/USPTO.spool_p/US09750986/runat_03022005_071743_18565/seq_query.faeta_1.775
-DB=GenEmbl -QFMT=fastap -SUFFIX=xrge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09750986@CGN1_1_5600@runat_03022005_071743_18565 -NCPU=6 -ICPU=3
-NO_WMAP -LANG=QUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WAPN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:.*
1: gb_da.*
2: gb_hcg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_srs.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	3082	100.0	1770	6 A92118 Sequence 29
2	3082	100.0	4166	1 A92118 Sequence 29
3	3082	100.0	32679	6 A92090 Sequence 1
4	2474	80.3	310029	1 A9016866 Pseudomon

5	2449	79.5	9477	1	PFL536325	AJ536325 Pseudomon
6	2349	76.2	301214	1	AB016786	AE016786 Pseudomon
7	2103	68.2	4407	1	PEU536324	AJ536324 Pseudomon
8	1940.5	63.0	189050	1	AL646077	AL646077 Ralstonia
9	1652.5	53.6	1902	6	AR319843	AR319843 Sequence
10	1621.5	52.6	45324	1	ACCEPCAC	L05770 Acinetobact
11	1618.5	47.5	110000	1	CR543861_17	Continuation (18 o
12	1457	42.3	346879	1	BS572598	BX572598 Rhodospir
13	1401	45.5	299950	1	AP005963	AP005963 Bradyrhiz
14	1294	42.0	10173	1	AE008067	AE008067 Agrobacte
15	1294	42.0	10173	1	AE009102	AE009102 Agrobacte
16	1210.5	39.3	299450	1	AP005938	AP005938 Bradyrhiz
17	1129	36.6	299950	1	AP005963	AP005963 Bradyrhiz
18	1002.5	32.5	10306	1	AE005712	AE005712 Caulobact
19	972	31.5	12606	1	AE005995	AE005995 Caulobact
20	799	25.9	2188	6	AX024541	AX024541 Sequence
21	766.5	24.9	2171	6	AX024537	AX024537 Sequence
22	743	24.1	7171	1	AF302797	AF302797 Pseudomon
23	698.5	22.7	1203	6	AX024534	AX024534 Sequence
24	337	10.9	301457	1	AE016924	AE016924 Chromobac
25	335.5	10.9	299925	1	AP005043	AP005043 Streptomy
26	325	10.5	10395	1	AE012877	AE012877 Chlorobiu
27	319.5	10.4	310550	1	SC0939113	AL939113 Streptomy
28	314.5	10.2	10502	1	AE014645	AE014645 Bifidobac
29	314.5	10.2	349980	6	AX492782	AX492782 Sequence
30	314.5	10.2	349980	6	AX553949	AX553949 Sequence
31	310.5	10.1	2109	5	BX936010	BX936010 Gallus ga
32	309	10.0	38494	6	AR345349	AR345349 Sequence
33	309	10.0	38503	1	MSGB1912CS	L01536 M. leprae g
34	302.5	9.8	35961	6	MLU5181	U15181 Mycobacteri
35	302.5	9.8	35961	6	AR345370	AR345370 Sequence
36	302.5	9.8	348450	1	MLBPRTN4	AL583920 Mycobacte
37	301.5	9.8	37286	6	CQ363734	CQ363734 Sequence
38	301.5	9.8	110000	1	AE017283_24	Continuation (25 o
39	294.5	9.6	11198	1	AE004318	AE004318 Vibrio ch
40	294.5	9.6	248550	1	SC0939120	AL939120 Streptomy
41	290	9.4	349944	1	BX571871	BX571871 Phototrab
42	289	9.4	10029	1	AE013966	AE013966 Versinia
43	289	9.4	201050	1	AJ414143	AJ414143 Versinia
44	289	9.4	293371	1	AE017141	AE017141 Yersinia
45	289	9.4	300900	1	AP005939	AP005939 Bradyrhiz

ALIGNMENTS

RESULT 1
A92118 LOCUS 1770 bp DNA linear PAT 22-JAN-2000
DEFINITION Sequence 29 from Patent EP0845532.
ACCESSION A92118
VERSION A92118.1 GI:6740914
KEYWORDS
SOURCE
ORGANISM
unidentified
unidentified
unclassified.
REFERENCE
1 (bases 1 to 1770)
Priester, H.D. and Rabenhorst, J.D.
Enzymes for the synthesis of coniferyl alcohol, coniferyl aldehyde,
ferulic acid, vanillin, vanillic acid and their applications
Patent: EP 0845532-A 29 03-JUN-1998;
HAARMANN & REIMER GMBH (DE)
FEATURES
source
1. 1770
/organism="unidentified"
/mol_type="unassigned DNA"
/db_xref="taxon:32644"
1. 1770
/gene="PCS"
1. 1770
/gene="PCS"
/codon_start=1
/product="FERULASACURE-COA-SYNTHETASE"
/protein_id="CAB69489.1"

```

/db xref="GI:6740915"
/translation="MRSLEBALPFGRIILERLEHMAKTRPEQTCVAAARANBEMRIS
VEMFNVAIAIOSILPGLSARPLIYSGNDLEHLQAFGMVAGIYCPVSPYS
ILSOLAKRHTVGLQPGIYVRAADAPROAIERTLPDVAIIFRGLAARTYSP
DSLREPGITEADNAPATATGPDTIAKFLTSGSTKLPKAVPTTORMLCANQMLDTF
PVFGEPPVLVWLPWNHTFGSGSHNIGIYLVNGTYLDDGKPTAGFAETLRNISEI
SPTAYLTVPKMBELVGLERDSTLEREPARKLFPFAAGLSQGIWRLRLVAEQH
CEERIRMAAGLMTETAPSCFTTGTGPAAGYGLAPPCGEVLVVDGKLEFRFGP
HYMSGYRAPEONAOAFDEGGYCSGDALKLADPADPOKGYMDGR IAEFDKLSGVF
VSGPLRTPAVLEGGSYVLDVVVAARECDICLIYVPRLLDQRAISGLGKEASDEVL
ASEPVRAMPADMLKILNREATGNASRIIMVGLIDTPPSIDKEVTDKSINQRAVLQM
RSKAVDALYRGDSMLRDEATL"

```

ORIGIN

Alignment Scores:

Pred. No.:	9,38e-211	Length:	1770
Score:	3082.00	Matches:	589
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-09-750-986d-30 (1-589) x A92118 (1-1770)

```

QY 1 MetArgSerLeuGluAlaLeuLeuProPheProGlyArgIleLeuGluuArgLeuGluHis 20
DB 1 ATGCCGTTCTCTCGAGGCGCTTCTTCCCTCCCGGGTCAATTCCTTGAGCGTCTCGAGCAT 60
QY 21 TTPAlaIyGthrArgProGluGlnThrCyValAlaAlaArgAlaAlaAengIyGluTrp 40
DB 61 TGGGCTMAAACCCGTCACAAACAACTCGTTCGCGACGAGCGGCAATGGGAATGG 120
QY 41 ArgArgIleSerTyrAlaGluMetPheHisAsnValArgAlaIleAlaGlnSerLeuLeu 60
DB 121 CCGTCGATACGCTACCGGAAATGTTCCACACGTCGCGGCATCCGACAGACTTGCTT 180
QY 61 ProTyrGlyLeuSerAlaGluArgProLeuLeuIleValSerGlyAsnAspLeuGluHis 80
DB 181 CCTTACGGACTATCGGACAGCGCTCGCTGCTTATCGTCTGGAATGACCTCGAATCAT 240
QY 81 LeuGlnLeuAlaPheGlyAlaMetTyrAlaGlyIleProTyrCyAspProValSerProAla 100
DB 241 CTTACGCTGGCAATTGGGGCTATGATGGGGCATTCCTATTGGCCGGGTCTCGCT 300
QY 101 TyrSerLeuLeuSerGlnAspLeuAlaIySLeuArgHisAlaValGlyLeuLeuGlnPro 120
DB 301 TATTCTGCTGTCGGAAGATTTGGCGAAGCTGCTGCACATCGTAGTCTTCTGCAACCG 360
QY 121 GlyLeuValPheAlaAlaAspAlaAlaProPheGlnArgAlaIleGluThrIleLeuPro 140
DB 361 GGACTGGTCTTGGCTCCGATGACGACCTTTCACGCGCAATTGAGACCATTTGCGCG 420
QY 141 AspAspValProAlaIlePheThrArgGlyGluLeuAlaGlyArgArgThrValSerPhe 160
DB 421 GACGAGTCCCGCCCAATCTTCACTCGAGGAAATTTGGCGGGCGGCGCAAGGATTTT 480
QY 161 AspSerLeuLeuGluGlnProGlyGlyIleGluAlaAspAsnAlaPheAlaAlaThrGly 180
DB 481 GACAGCTCTCGAGAGAGCTGGTGGATTGAGGCAAGATATGCCCTTTGCGGCACTGGC 540
QY 181 ProAspThrIleAlaIySLeuPheLeuPheThrSerGlySerThrIySLeuProIySAlaVal 200
DB 541 CCCGATACGATGGCCAAATCTTGTTCACCTTCTGCTTACCAAACTGCTTGAAGCGG 600
QY 201 ProThrThrGlnArgMetLeuCyAlaAengGlnGlnMetLeuLeuGlnThrPheProVal 220
DB 601 CCGACTACTCAGGGAATGCTTGGCGCAATCAGACAAATGCTTCTGAAACTTTCCGGGT 660
QY 221 PheGlyGlyGluProProValLeuValaAspTrpLeuProTrpAsnHisThrPheGlyGly 240
DB 661 TTGGGGAAGAGCGCGGCTGCTGAGTGTGATTCCTGGAACACACACTTTCGCGCG 720
QY 241 SerHisAsnIleGlyIleValLeuTyrAsnGlyGlyThrTyrTrpLeuAspAspGlyLys 260

```

```

DB 721 ACCCAACAATCGGATCGTGTGTGACAAAGCGGCGCATGACTACCTTACACACGCTAA 780
QY 261 ProThrAlaGlnGlyPheAlaGluThrLeuArgAsnLeuSerGluIleSerProThrAla 280
DB 781 CCAACCGCCCAAGGGTCCCCCAAGCGCTTCCCACTTGAAGGAATCTCTCCCATGGG 840
QY 281 TyrLeuThrValProIySgIyTrpGluGluLeuValGlyAlaLeuGluuArgAspSerThr 300
DB 841 TACCTCACTGTCGCCGAAGAGCTGGAGGAATAGTGGTGCCCTTGAGGAGACAGTACC 900
QY 301 LeuArgGluArgPhePheAlaArgMetLysLeuPhePhePheAlaAlaAlaGlyLeuSer 320
DB 901 CTGCGCGAAAGCGCTTCTTCCCTCGCATGAAGCGTTCCTTCTGCGCGCGCTGGGTGTG 960
QY 321 GlnGlyIleTrpAspAspArgLeuAspArgValAlaGluGlnHisCyAspGlyGluArgIleArg 340
DB 961 CAAGGATCTGGGATCGTGTGACCGGGTGGCTGAACACAGCACTGTGTAGGCGCATTTCC 1020
QY 341 MetMetAlaGlyLeuGlyMetThrGluThrAlaProSerCyThrPheThrThrGlyPro 360
DB 1021 ATGATGGCGGGTCTGGGCATGACGAGACTGCTCTTCTGCACTTTTACACCGGACCG 1080
QY 361 LeuSerMetAlaGlyTyrIleGlyLeuProAlaProGlyCyGluValLysLeuValPro 380
DB 1081 CTGTGATGGCTGGTTACATGGGCTGCCAGCGCTGGCTGGAGGTCAAGCTCGTTCCG 1140
QY 381 ValAspGlyLysLeuGluGlyArgPheHisGlyProHisValMetSerGlyTyrTrpArg 400
DB 1141 GTGAGTGGAAATTTGGAAGGGGGTTCATGGTCCGACCTCATAGCGGCTACTGGCT 1200
QY 401 AlaProGluGlnAsnAlaGlnAlaPheAspGluGluGlyTyrTyrCySerGlyAspAla 420
DB 1201 GCTCCGAAACAATAAGGCCCAAGCGTTCCAGCGAAGGCTTATTCGCTCCGGTATGCC 1260
QY 421 IleLysLeuAlaAspProAlaAspProGluGlyGlyLeuMetPheAspGlyArgIleAla 440
DB 1261 ATCAAAATGGCAGATCTCGCCATCTCGAAAGGCTGATGTTGACGCTGGAATGGT 1320
QY 441 GluAspPheLysLeuSerSerGlyValPheValSerValGlyProLeuArgThrArgAla 460
DB 1321 GAAAGCTTAAAGCTGTCTTCAAGGGATTATTTGCAAGGCTTGGGCCATTTGCGACGCGGGC 1380
QY 461 ValLeuGluGlySerTyrValLeuAspValAlaAlaAlaProAspArgGluCyS 480
DB 1381 GTTTCGAAAGGGCGGCTTACGTCCTGACGTAAGTGTGTGCTGCTCATCGTGAATGC 1440
QY 481 LeuGlyLeuLeuValPheProArgLeuLeuAspCyAspAlaLeuSerGlyLeuGlyLys 500
DB 1441 CTTGGAATGCTGTGTTCCTCGGGTCTTTCGACTGCGCGGCTTGTGCGGGGCTAAGAAAA 1500
QY 501 GluAlaSerAspAlaGluValLeuAlaSerGluProValArgAlaTrpPheAlaAspTrp 520
DB 1501 GAGGCGTCCGACGCCGAGGCTTGGCACTGACGCGGCTTGGGCTGCTGCTGACTGG 1560
QY 521 LeuLysArgLeuAsnArgGluAlaThrGlyAsnAlaSerArgIleMetTrpValGlyLeu 540
DB 1561 CTCAAACGACTCAATCGAACAACAAGCTGGCAATGCGATCATGTGGGTAGGGCTTC 1620
QY 541 LeuAspThrProProSerIleAspLysGlyGluValIleThrAspLysGlySerIleAengGln 560
DB 1621 CTCGATACCGCCCGCTGATTATTAAGGCGAGGTCACTGCAAGAGGCTCGATCAACCG 1680
QY 561 ArgAlaValLeuGluThrArgSerAlaLysValaAspAlaLeuTyrArgGlyGluAspGln 580
DB 1681 CGGCTGTTTTCATATGGCGGTGGGGAAGTTGATGCCCTTATTCGTGTGAATCAA 1740
QY 581 SerMetLeuArgAspGluAlaThrLeu 589
DB 1741 TCCATGCTGCGTACGAGGCGCACACTG 1767

```

RESULT 2
PSP238746

LOCUS PSP238746 4166 bp DNA linear BCT 05-NOV-1999
DEFINITION Pseudomonas sp. vdh (partial), fcs, aat and mac (partial) genes,
strain HR199.
ACCESSION AJ238746
VERSION AJ238746.1 GI:6273620
KEYWORDS aat gene; beta-ketochiolase; fcs gene; feruloyl-CoA synthetase; mac
gene; methyl-accepting chemotaxis protein; vanillin dehydrogenase;
vdh gene.
SOURCE Pseudomonas sp.
ORGANISM Pseudomonas sp.
Bacteria; Proteobacteria.
REFERENCE
1 Overhage, J., Priefert, H. and Steinbuechel, A.
Biochemical and genetic analyses of ferulic acid catabolism in
Pseudomonas sp. Strain HR199
Appl. Environ. Microbiol. 65 (11), 4837-4847 (1999)
JOURNAL MEDLINE 20011220
PUBMED 10543794
2 (bases 1 to 4166)
Priefert, H.
Direct Submission
Submitted (04-MAY-1999) Priefert H., Westfaelische
Wilhelms-Universitaet Muenster, Institut fuer Mikrobiologie,
Corrensstrasse 3, Muenster, D-48149, GERMANY
FEATURES
source
1. .4166
/organism="Pseudomonas sp."
/mol_type="genomic DNA"
/strain="HR199"
/db_xref="taxon:306"
/country="Indonesia"
1. .170
/gene="vdh"
<1. .128
/gene="vdh"
/function="NAD dependent oxidation of vanillin to vanillic
acid"
/codon_start=3
/evidence=experimental
/transl_table=11
/product="vanillin dehydrogenase"
/protein_id="CAB60225.1"
/db_xref="GI:6273621"
/db_xref="GOA:Q9RLD0"
/transl_table=11
/translation="QMPFGVSSVSGSGSRASIEHTQLRLTITQNGRHYPI"
138. .170
/gene="vdh"
/function="putative terminator"
join(373..378,388..2157)
/gene="fcs"
373..378
/gene="fcs"
388..2157
/gene="fcs"
/function="ATP dependent conversion of ferulic acid to
feruloyl-Coenzyme A"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="feruloyl-CoA synthetase"
/protein_id="CAB60226.1"
/db_xref="GI:6273622"
/db_xref="GOA:Q9RLD9"
/db_xref="TrEMBL:Q9RLD9"
/translation="MRSLEALLPFGRIERLEHMAKTRPEQTCVAAARANGEMRIS
VAMFNVNVAIAOSLIPYGSARPLIYSGNDLEHLOLAFGMYAGIPLYCPSPAYS
LISODLAKLRHIVGLDPGLVFAADAAPRRALETILPDVPAIFRGLAGARTYSP
DSLEPGGIBADNAPARATSPDIKFLPFGSTKPKAVPTTQRNLCAHQMLDTF
PVFGEPPVLVDWLPNNHTFGSHNIGIVYNGTYYLDGKFTAGSFATLNLSEI
SPRAYLTVPKGMEVLGALERDSTLEREPARKMLFFPAAGISQGIWRLDVAQOH
CGRRIMMAGLGTETAPSCFTTGPGLSMAGYGLPAPGCEVLIVDVGKLEGRFPG
HMSGVYARAEQVNAQAFDEGGYCSGDATKLDADPADPQGLMFGRIADBFKLSGVF

VSVPILRTAVLEGSSVLDVVVAADRECLGLVFPRLDLCALSGLEKASDAEVL
ASPPVAMFADMLKRLNREATGNASIMVVGILDDTPESIDKSVTKGSINGRAVLQW
RAKVDALYRGEDQSLRDEATL"
gene
2143..2146
/gene="aat"
RBS 2143..2146
/gene="aat"
gene 2154..3449
/gene="aat"
2154..3449
/gene="aat"
CDS
/codon_start=1
/transl_table=11
/product="putative beta-ketochiolase"
/protein_id="CAB60227.1"
/db_xref="GI:6273623"
/db_xref="TrEMBL:Q9RLD8"
/translation="MSWSGGAVSAPSDTLVNAVTRPTWIDCGGALSVPIDLVKVA
REVLMRASLEPQWVDVSLAGMAQASFDVLLPRHIGLYSGVPSVPALGVORICGTG
FELLROGROIQSGADHYLCVAESMSRPDIASVYTRGGFRGAPVEPDFLMEALFD
PARGDMINTAERLARLYGITTREANSYAVSSPERALPAOEKWIQOETVYATDQED
LGGYNRAIELEPKAKLLIVYTRGLAVEALSRLKPVHSGVQTGNSCAVVDGAAA
ALVARSSTOPIVRLIILATSVGLIEPHMGGLGAPARILRLARSDLSLRDIDLFEIN
EAQAQVLAVOHELGIEHSKLNIMGALIGHPLAATGRLCWTTLAHOLQANNFRYGI
ASACIGGGQGMVLTLENPHFGSSSARSMINRDVPLS"
3521..4166
/gene="mac"
3521..3525
/gene="mac"
CDS 3535..4166
/gene="mac"
/codon_start=1
/transl_table=11
/product="putative methyl-accepting chemotaxis protein"
/protein_id="CAB60228.1"
/db_xref="GI:6273624"
/db_xref="TrEMBL:Q9RLD7"
/translation="MISFAPMAESLGVQAKLALFALVLCVGLIVNGTGFVHTLSG
LVKSAIAGEIRAKTOELKVLQORALFIADGSLKORSLISQVLAENVDALIDIPRO
RGSEILKFRASREASYSIEVGSNAADKLGSGESDLMADKKNLYEEDLSANV
NALMLGLIEDONEKPLIYIMLGVTTLFMLSAYSVPWFSRQLVPELAKST"
ORIGIN
Alignment Scores:
Pred. No.: 2,86e-210 Length: 4166
Score: 3082.00 Matches: 589
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0
US-09-750-986D-30 (1-589) x PSP238746 (1-4166)
Qy 1 MetArgSerLeuGluAlaLeuLeuProPheProGlyArgGlyLeuGluAlaLeu 20
Db 388 ATCGCTTCTCTCGAGCGCTTCTTCCCTTCCCGGCTCGAATTTTAGAGCTCTCGAGCAT 447
Qy 21 TTPAlaLySThArgProGluGlnThrCySValAlaAlaArgAlaAlaLeuGlyGluTTP 40
Db 448 TGGGCTAAGCCCGCTCAGAACAAACCTGGCTGTCGCCAGCGCGCAATGGGGAATCG 507
Qy 41 ArgArgTLLSerTyrTrpArgIuMetPheHISAsnValArgAlaAlaLeuArgInsSerLeuLeu 60
Db 508 CGTGTATCAGCTACCGGAAAGTTTCCACAAAGCTCCGCGCCATTCGCACAGACTTGCTT 567
Qy 61 ProTyrGlyLeuSerZAlaGluArgProLeuLeuIleValSerGlyVanAspLeuGluHIS 80
Db 568 CTTTAGCGACTATCGGACAGAGCGTCCGCTTATGCTCTCGGAATGACCTCGGAACAT 627
Qy 81 LeuGluLeuAlaPheGlyAlaMetTyrAlaGlyTLeProTyrCySProValSerProAla 100
Db 628 CTTACAGCTGCAATTTGGGGCTATGATGCGGGGCAATTCCTAATGGCCCGGTCTCTGCT 687

QY 101 TyrSerLeuLeuSerGlnAspLeuAlaValLeuArgHisIleValGlyLeuLeuGlnPro 120
 DB 688 TATTCACTGCTGCGCAAGATTGGCGAAGCTCGTCACATCTAGGCTTTCTGCAACCG 747
 QY 121 GlyLeuValPheAlaIleAspAlaAlaProPheGlnArgAlaIleGlnThrIleLeuPro 140
 DB 748 GGACTGGCTCTTGCTCCGATGACGACCTTTCACGCGCAATTGACATCTTCGCG 807
 QY 141 AspAspValProAlaIlePheThrArgGlyGluLeuAlaGlyArgArgThrValSerPhe 160
 DB 808 GAGAGGTGCGCCGCAATCTTCACTCGAGCGAATTGGCCGGCGCGCAAGGTGATTT 867
 QY 161 AspSerLeuLeuGlnGlnProGlyGlyIleGluAlaAspAsnAlaPheAlaAlaThrGly 180
 DB 868 GACAGCTCTGTCGACGACCTGGTGGGATTGAGCGAGATATCTTTGGCGCACTGGC 927
 QY 181 ProAspThrIleAlaIlePheLeuPheThrSerGlySerThrIleLeuProValAlaVal 200
 DB 928 CCGATACGATTGCCAAGTCTTGTCATCTTGGCTTACCAACTGCTTAAGSGCGT 987
 QY 201 ProThrThrGlnArgMetLeuCysAlaAsnGlnGlnMetLeuLeuGlnThrPheProVal 220
 DB 988 CGACCTACTCAGGAGATGCTCTGCGCCATCAGCAGATGCTTTCGCAACTTCCGGTT 1047
 QY 221 PheGlyGluGluProProValLeuValAspTTPLeuProTTPAsnHisThrPheGlyGly 240
 DB 1048 TTTGGTGAAGAGCCGCGGTGCTGGTGAAGTCTTCCGTGGAACCACTTCCGCGCG 1107
 QY 241 SerHisAsnIleGlyIleValLeuTyrAsnGlyGlyThrTyrThrLeuAspAspGlyVal 260
 DB 1108 AGCACAACATCGGCATCGGTGTACAAACGCGGCACTGACTTGAACGCGGTAA 1167
 QY 261 ProThrAlaGlnGlyPheAlaGluThrLeuArgAsnLeuSerGluLeuSerProThrAla 280
 DB 1168 CCAACGCCCAAGGGTTCGCGGAGAGCTTTCGCACTTGAAGCAATCTTCCACTGCG 1227
 QY 281 TyrLeuThrValProIleGlyTTPGluGluLeuValGlyAlaLeuGluArgAspSerThr 300
 DB 1228 TACCTCACTGTCGGAAGGCTGGGAGGAAATAGTGGGTCCCTTGAAGCGAGCACTAC 1287
 QY 301 LeuArgGluArgPhePheAlaArgMetLeuPhePhePheAlaIleGlyLeuSer 320
 DB 1288 CTGGCGGAACGCTTCTTCCGTGCAATGAGCTGTCTTCTCGCGCGGTGGGTGTGCG 1347
 QY 321 GlnGlyIleTTPAspArgLeuAspArgValAlaGluGlnHisCysGlyGluArgIleArg 340
 DB 1348 CAAGGATCTGGGATCGTTTGACCGGGTCCGGAACAGACGTGTGAGCGCATTTCCG 1407
 QY 341 MetMetAlaGlyLeuGlyMetThrGluThrAlaProSerCysThrPheThrThrGlyPro 360
 DB 1408 ATGATGCGCGGTCTGGGCAATGACGAGACTGCTCTTCCGCACTTTTACACCGGACCG 1467
 QY 361 LeuSerMetAlaGlyTyrIleGlyLeuProAlaProGlyCysGluValIleValLeuValPro 380
 DB 1468 CTGTGATGCTGGTTTACATTTGGCTTCCAGCGCTGCGAGGTCAAGCTCGTTCCG 1527
 QY 381 ValAspGlyValLeuLeuGluArgPheHisGlyProHisValMetSerGlyTyrTTPArg 400
 DB 1528 GTCGATGGGAAATGGAGGGGCTTTCATGCTCCGACGTCATGAGCGGCTACTGGCGT 1587
 QY 401 AlaProGluGlnAsnAlaGlnAlaPheAspGluGluGlyTyrTyrCysSerGlyAspAla 420
 DB 1588 GCTCCCTGAACAAATGCCAAGGCTTCGACGAGGAGGCTATTATCTCCGCTGATGTC 1647
 QY 421 IleValLeuAlaAspProAlaAspProGlnIleValLeuMetPheAspGlyArgIleAla 440
 DB 1648 ATCAATTTGGCAGATCTGCGCATCTCAAGAAAGGCTGATGTTTACAGGTCAATTTGCT 1707
 QY 441 GluAspPheValLeuSerSerGlyValPheValSerValGlyProLeuArgThrArgAla 460
 DB 1708 GAAGACTTCAGGTGCTCAAGGCTATTGTTCAGGCTTGGGCGCATTTGGCGACGCGGG 1767
 QY 461 ValLeuGluGlyGlySerTyrValLeuAspValValAlaIleAlaProAspArgGluCys 480

DB 1768 GTTCTGGAAGGGCGCTTTCAGTCTGAGCAGTGTGCTGCTCTTATCTGTAATGC 1827
 QY 481 LeuGlyLeuLeuValPheProArgLeuLeuAspCysArgAlaLeuSerGlyLeuGlyVal 500
 DB 1828 CTGGATGCTGTGTGTTTCCGGGTCTTCTGACTGCGGTCTTTCGCGGCGTGAAGAAA 1887
 QY 501 GluAlaSerAspAlaGluValLeuAlaSerGluProValArgAlaIlePheAlaAspTTP 520
 DB 1888 GAGCGCTCGACCGCCAGGTGCTTGCAGTGAAGCCGCTTCCGGCTGTTTCTGACTGG 1947
 QY 521 LeuValArgLeuAsnArgGluAlaThrGlyAsnAlaSerThrGlyMetTTPValGlyLeu 540
 DB 1948 CTCMAACGACTCATCTGAAAGCAACTGCAATGCGCATCTGTGGGTAGGGCTTC 2007
 QY 541 LeuAspThrProProSerIleAspIleValIleValThrAspIleGlySerIleAsnGln 560
 DB 2008 CTGATACCGCCCGCTGATTTGATTAAGGCGAGGTACTGACAAAGGCTCGATCAACCG 2067
 QY 561 ArgAlaValLeuGlnTTPArgSerAlaValAspAlaLeuTyrArgGlyGluAspGln 580
 DB 2068 CGGCTGTTTTCATGCGGTGCGGGAAGTTGATGCCCTGTATGTGTGTAAGATCA 2127
 QY 581 SerMetLeuArgAspGluAlaThrLeu 589
 DB 2128 TCCATGCTGCTGACGAGGCCACACTG 2154

RESULT 3

LOCUS A92090 32679 bp DNA linear PAT 22-JAN-2000
 DEFINITION Sequence 1 from Patent EP0845532.
 ACCESSION A92090
 VERSION A92090.1 GI:6740886
 KEYWORDS
 SOURCE Pseudomonas sp.
 ORGANISM Pseudomonas sp.
 Bacteria; Proteobacteria.
 REFERENCE
 1 (bases 1 to 32679)
 Pfeifer, H.D. and Rabenhorst, J.D.
 Enzymes for the synthesis of conferyl alcohol, conferyl aldehyde,
 fennilic acid, vanillin, vanillic acid and their applications
 Patent: EP 0845532-A 1 03-JUN-1998;
 HAAWMANN & REIMER GMBH (DE)
 JOURNAL Location/Qualifiers

FEATURES

source 1..32679
 /organism="Pseudomonas sp."
 /mol_type="unassigned DNA"
 /strain="HR199"
 /db_xref="taxon:306"
 gene 3146..4000
 /gene="ORF1"
 CDS 3146..4000
 /gene="ORF1"
 /codon_start=1
 /transl_table=11
 /protein_id="CAB69475.1"
 /db_xref="GI:6740887"
 /translation="MIAITGASGQGRITLALRLKRPASRIIALVRDKNKAGDLTAR
 GIVRQADYNREPETHRLALIGVNRLLIISSEVGORTQHRVITDAKQEGIELAYT
 SILHDKSNLGLATHRDTEQALTESGIPHYTLRQGWENTYTAGIPVALVGVLLC
 AODGILASARADYVAPALAVITGENQKRVTELNGEPATYTTTEIAAEVAPQAGTIV
 YSNLSESDYRSRLISAGLPDGPALLADSDAGAAKGYLFDSSGDSRKLIGRPTTPE
 AIAAAIGR"

ORIGIN

Alignment Scores:
 Pred. No.: 4,2e-209 Length: 32679
 Score: 3082.00 Matches: 589
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

us-09-750-986d-30 (1-589) x A92090 (1-32679)

QY 1 MetArgSerLeuGluAlaLeuLeuProPheProGlyValArg11LeuGluAlaArgLeuGluHis 20
 Db ATGGCTCTCTCCAGAGCGCTTCTTCCCTCCGCGGCTGAATCTTGAGCCCTCCAGCAT 21347

QY 21 TPpAlaLysThrArgProGluGluInThrCysValAlaAlaArgAlaAlaAsnGlyGluTrp 40
 Db TGGGCTTAAGACCCGCTCCAGAACAACTGGGTTGCTCCAGGGCGGCAATGGGAAATGG 21407

QY 41 ArgArg11SerTrpAlaGluMetPheHisAsnValArgAla11LeuArg11SerLeuLeu 60
 Db CGTGGATACGATACGGGAAATGTTCCACACGTCGGCCCATCGCACAGAGCTTGCTT 21467

QY 61 ProTrpGlyLeuSerAlaGluArgProLeuLeu11Val1SerGlyAsnAspLeuGluHis 80
 Db CCTTACGGAATACGGAGAGCGCTCCGCTTATCTCTCTGAAATGACCTGGAAATCAT 21527

QY 81 LeuGluLeuAlaPheGlyAlaMetTrpAlaGly11LeuProTrpCysProVal1SerProAla 100
 Db CTTCAGCTGGCAATTTGGGCTTAATGATGCGGCAATTCCTAATGCCCCGATGCTCTGCT 21587

QY 101 TrpSerLeuLeuSerGlnAspLeuAlaLysLeuArgHis11LeuAlaGlyLeuLeuGlnPro 120
 Db TATTCACTGCTGCGCAAGATTTGGCGAGACTGGTCACATCGAGGCTTCTGGCAACCG 21647

QY 121 GlyLeuVal1PheAlaAlaAspAlaAlaProPheGlnArgAla11LeuTrp11LeuPro 140
 Db GGACTGCTCTTCTCCGATGACGACCTTTCACGCGCCCAATGAGACCATTCGCGC 21707

QY 141 AspAspVal1ProAla11PheThrArgGlyGluLeuAlaGlyArgArgTrpVal1SerPhe 160
 Db GAGGAGGTGGCCCCCAATCTTCACTCCAGGCAATGGCCGCGGCGCCAGAGTGGT 21767

QY 161 AspSerLeuLeuGluGlnProGlyGly11LeuAlaAspAsnAlaPheAlaAla11ArgGly 180
 Db GACAGCTGCTGGAGACGCTGGGGGATTGAGGCAATATGCTTGGCGCAACTGGC 21827

QY 181 ProAspThr11LeuAlaLysPheLeuPheThrSerGlySerThrLysLeuProLysAlaVal 200
 Db CCCGATACGATTCGCAAGTCTTGTTCATCTGGGCTTACCAAACTGGCTTAAGCGCGTG 21887

QY 201 ProThrThrGlnArgMetLeuCysAlaAsnGlnGluMetLeuLeuGlnThrPheProVal 220
 Db CCAGACTACTCAGCAATGCTCTGGCCATACAGAGATGCTTCTGCAACTTTCGCCGCT 21947

QY 2188 PheGlyGluGluProProVal1LeuVal1AspTrpLeuProTrpAsnHisThrPheGlyGly 240
 Db TTTGGTGAAGAGCGCGCGCTGGTGGACTGGTGGCGTGGAAACCAACCTTCGCGCGC 22007

QY 241 SerHisAsn11LeuGly11LeuVal1LeuTrpAsnGlyGlyThrTrpLeuAspAspGlyLys 260
 Db AGCCACAACATCGCATCGTGTGTACACAGCGGCGCACGTAATCACTTGACGACGCTAAA 22067

QY 261 ProThrAlaGlnGlyPheAlaGluTrpLeuArgAsnLeuSerGlu11SerProThrAla 280
 Db CCAACCGCCCAAGGGTTCGCCGAGACGCTTGGCACTTGAGCAAAATCTTCCACTGCG 22127

QY 281 TrpLeuThrVal1ProLysGlyTrpGluGluLeuVal1AlaLeuGluArgAspSerThr 300
 Db TACCTACTGCTGGCGAAAGCTGGAGGATTAATGTTGGTCCCTTGAGCGAGACAGTACC 22187

QY 301 LeuArgGluArgPhePheAlaArgMetLysLeuPhePheAlaAlaAlaGlyLeuSer 320
 Db CTCGCGGACAGCTTCTTCCCTCCAGTACACTGTTCTTCCGCGGCGCTGGGTTGTCG 22247

QY 321 GlnGly11LeuTrpAspArgLeuAspArgVal1AlaGluGlnHisCysGlyGlyArg 340
 Db CAAAGGATCTGGGATCTTGGACCGGGTGGCTGAAGACATGTTGGTGAAGCGCATTCGC 22307

QY 341 MetMetAlaGlyLeuGlyMetTrpGluTrpAlaProSerCysThrPheThrArgLysPro 360
 Db ATGATGGCGGCTCTGGGCAATGAGAGACTGCTCTTCCGCACTTTTACACCGAGCGG 22367

QY 361 LeuSerMetAlaGlyTrp11LeuGlyLeuProAlaProGlyCysGluValLysLeuValPro 380
 Db CTGTCAATGGCTGTATACATTTGGGCTGGCCAGCCCTGGCTGGCAAGCTCTGTCG 22427

QY 381 ValAspGlyLysLeuGluGlyArgPheHisGlyProHis11ValMetSerGlyTrpTrpArg 400
 Db GTTCATGGGAAATTTGAAAGGCGCTTTCATGCTCCGCACTCATGAGCGGCTACTGGCGT 22487

QY 401 AlaProGluGlnAsnAlaGlnAlaPheAspGluGlyGlyTrpTrpCysSerGlyAspAla 420
 Db GCTCTGAACAAATATCCCAAGCGTTTCSACGAGAAAGGCTATATGCTCTCCGCTGATGC 22547

QY 421 11LeuLeuAlaAspProAlaAspProGluLysGlyLeuMetPheAspGlyArg11LeuAla 440
 Db ATCAAAATGGCAATCCCTGCGATCCCTCAAGAAAGTGTGATGTTGAGGTCGAATGCT 22607

QY 441 GluAspPheLysLeuSerSerGlyVal1PheValSerValGlyProLeuArgTrpArgAla 460
 Db GAAGACTTCAAGCTGCTCTCAGGGGATTTGTCAAGCGTTGGCCATTGGCCACGCGCGG 22667

QY 461 ValLeuGluGlyGlySerTrpVal1LeuAspVal1ValAlaAlaProAspArgLysCys 480
 Db GTTCTGAAGGCGGCTCTTACGTCTGACGTAGCTGCTGCTCTCTCTGATGTC 22727

QY 481 LeuGlyLeuLeuVal1PheProArgLeuLeuAspCysArgAlaLeuSerGlyLeuGlyLys 500
 Db CTTGATGCTCTGTGTTCCGCTCTTCTGACTGCGGCTTGCTGCGGGCTAGGAAA 22787

QY 501 GluAlaSerAspAlaGluVal1LeuAlaSerGluProValArgAlaTrpPheAlaAspTrp 520
 Db GAGGCGTGGAGCCCGAGAGCTTGGCAGTACGACCGGTTGGGCTGCTGACTGG 22847

QY 521 LeuLysArgLeuAsnArgGluAlaTrpGlyAsnAlaSerArg11MetTrpVal1GlyLeu 540
 Db CTCAAACGACTCAATGAGAGCACTGGCAATGCCAGTGCATCATGTGGTGAAGGCTC 22907

QY 541 LeuAspThrProProSer11LeuLysGlyGluVal1ThrAspLysGlySer11LeuGln 560
 Db CTGATACGCGCGCGTGTGATTAAGGGGAGAGTCACTGACAAAGGCTGATCAACAG 22967

QY 561 ArgAlaValLeuGlnTrpArgSerAlaLysValAspAlaLeuTrpArgGlyGluAspGln 580
 Db CGGCGTGTGTTGCAATGGCGGTGGCGAAAGTTGATCGCTGTATGCTGTGAAGTCAA 23027

QY 581 SerMetLeuArgAspGluAlaThrLeu 589
 Db TCCATGCTGCGTGAAGGCGCACACTG 23054

RESULT 4
 AE016866/c 310029 bp DNA linear BCT 09-OCT-2003
 LOCUS Pseudomonas syringae pv. tomato str. DC3000 section 11 of 21 of the complete genome.
 DEFINITION
 ACCESSION
 AE016866 AE016853
 VERSION
 AE016866.1 GI:28853240
 KEYWORDS
 SOURCE
 ORGANISM
 Pseudomonas syringae pv. tomato str. DC3000
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 Pseudomonadaceae; Pseudomonas.
 1 (bases 1 to 310029)
 REFERENCES
 AUTHORS
 Buell,C.R., Joardar,V., Lindeberg,M., Selengut,D., Paulsen,I.T.,
 Gwinn,M.L., Dodson,R.J., Deboy,R.T., Durkin,A.S., Kolonay,J.F.,
 Madupu,R., Daugherty,S., Brinkac,L., Beaman,M.J., Haft,D.H.,
 Nelson,W.C., DavidSEN,T., Zafar,N., Zhou,L., Liu,J., Yuan,Q.,
 Khouri,H., Fedorova,N., Tran,B., Russell,D., Berry,K.,
 Kouturack,T., Van Aken,S.E., Feldblyum,T.V., D'Ascenzo,M.,
 Deng,W.L., Ramos,A.R., Alfaro,U.R., Cartimoun,S., Chatterjee,A.K.,
 Delaney,T.P., Lazarowitz,S.G., Martin,G.B., Schneider,D.J.,
 Tang,X., Bender,C.L., White,O., Fraser,C.M. and Collier,A.
 The complete genome sequence of the Arabidopsis and tomato pathogen
 Pseudomonas syringae pv. tomato DC3000

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 100 (18), 10181-10186 (2003)
PUBMED 12928499
REFERENCE 2 (bases 1 to 310029)
AUTHORS Berry, R., Joradar, V., Khouri, H., Fedorova, N., Tran, B., Russell, D.,
Berry, R., Uterback, T., Van Aken, S., Feldblum, T., Gwin, M.,
Nelson, W., Davidson, T., White, O., Fraser, C. and Collier, A.
TITLE Submitted (03-MAR-2003) The Institute for Genomic Research, 9712
JOURNAL Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source
1. .310029
/organism="Pseudomonas syringae pv. tomato str. DC3000"
/mol_type="genomic DNA"
/strain="DC3000"
/db_xref="taxon:223283"
/note="pathovar: tomato"
204. .1664
/locus_tag="PSPTO2809"
/locus_tag="PSPTO2809"
204. .1664
/locus_tag="PSPTO2809"
/note="similar to GP:15140142; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA056309.1"
/db_xref="GI:28853241"
/translation="MSDVNAAGLALAEARLODLMWLPAAKPKPPNAGAVTD
VAIGGMAGLALAEARLHGHGAAVYFDSPAFEPMPATTAMETLRSKQDTGAL
GPALEFRAMTEAQFGADGWAALDKIPLQWALYLRKALADVRNEHRSRVAR
TDGLVLDLMTPGQTRFMLARHYVALTGRGLGPPWPDFARQLPHLWTHSAAGD
DMEGRGAVVITGGGASANDSAATALAGAQVLLIRALPYNKKGKGVNGVGD
YMRPLPMKRIIRHYLNROQVPPRGSTLRSRSGARPLDNPVLAVERNPAAGVWL
QNSKARVADPVVATNGRFDPELRPEPAPESHIRVMODRFPAPGEGDSBLAVPD
LNCSPFOEKTTPACPGNLNHFSTPALISTAGVSGDIPATSEGSRLAHALGLLF
NBDIALHFSMSRYAEPLIGDEWVASEPTELRP"
1661. .2635
/locus_tag="PSPTO2810"
/locus_tag="PSPTO2810"
1661. .2635
/locus_tag="PSPTO2810"
/codon_start=1
/transl_table=11
/product="peptide ABC transporter, permease protein"
/protein_id="AA056310.1"
/db_xref="GI:28853242"
/translation="MTGMLRRLTOSLIVILMTVVVFGVNAIGNMDIIVGEDINO
ABRLQIADLGDQPLMOQYLIFLKGAMGNLQGSVYHEDAMRLILQRLPATFELAF
SAFLVAVIGVPLMGAGMTPDHPLSRLLMAASTVGFSLPAFVALMMLPSTITLQW
LPLSGGETREWLGVQSWLTLDDGLQLLLPALMLALFKISLVLRTLRAGVREVLQOE
FVYFARAKGLSPWRVCMHVMRTMTIPVTVLAMELGSITAYAVTESIFAMPGAGKL
ILDSINMLDRPVVAVLMTVVVIFVVLNLIVDGLYYLLDPRVIEASR"
2632. .3552
/locus_tag="PSPTO2811"
/locus_tag="PSPTO2811"
2632. .3552
/locus_tag="PSPTO2811"
/codon_start=1
/transl_table=11
/product="peptide ABC transporter, permease protein"
/protein_id="AA056311.1"
/db_xref="GI:28853243"
/translation="NNVIAKVAIPANTLLQAQSPWRVIVDFMSSSTAVVGLLLVIV
LRAAPAVIVQVNPDIQMLQNVDMADPGSLMDSGTYWLTGQDGLVSAIYIG
LRLSLVWIGSALIAVILGILGVAAGVMDALMLRLDILLSFPVITWMLMLTA
WCKGVGNMMLTILLEMAYVARTAGCATSRBRVYDARQGVGPRIYVGHILP
NCLPPIYVIGALQIARATLTBATLTSLGAGVPTESPSTGLLIANGQYMLSNYSWISL
FGLALITIVAINLVGRLRDVLPRLQR"
3549. .4562
/locus_tag="PSPTO2812"
/locus_tag="PSPTO2812"
3549. .4562
/locus_tag="PSPTO2812"
/codon_start=1

gene
/transl_table=11
/product="peptide ABC transporter, ATP-binding protein"
/protein_id="AA056312.1"
/db_xref="GI:28853244"
/translation="MSLPVSRVTSVEPTELVRLDCTSPYRAGVLPAVRNYSLVRO
PGRIITLVESSGSKSVTFTSLGIVDERGRISGGCVLKKRGLDTLSGQLEIGN
RVAHMFQDDMTMLPVLADTQMLTEAVRSHSTLGRKEAEHARSRTIALMGLISPERL
RAMPHQDSGMQORVAIAIALHAPDLIIADEPTTALDVTIQAQIISVQKIVREGQT
SLIWIHDSVVAGLADVDVAVVAGRIVEQSVADVLDVDPHPTQGLIDSLPSRMR
GQRLROIPEGMADLLSMPAGCAFAARCAKAPACQDPRPREILPGHTIRCFYAGCA
DNO"
4552. .5565
/locus_tag="PSPTO2813"
/note="This region contains an authentic point mutation,
causing a premature stop, and is not the result of a
sequencing artifact; peptide ABC transporter, ATP-binding
protein, authentic point mutation; similar to GP:15158466;
identified by sequence similarity; putative"
5618. .7183
/locus_tag="PSPTO2814"
/locus_tag="PSPTO2814"
5618. .7183
/locus_tag="PSPTO2814"
/note="similar to GP:15140494, and SP:P23847; identified
by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="peptide ABC transporter, periplasmic
peptide-binding protein"
/protein_id="AA056313.1"
/db_xref="GI:28853245"
/translation="MKPLVRSLLASALVLAQSVSAQDLRIGYADPVSSLDPLQNLNYA
GRRSVALHFEZLVSRDRLKPLGLAKSKVITDDTWELAEEDVEMQGTPLTADL
VSVFERARVSVAAGVAGAMRTVSVKAKDHTLIIKRLPNANLIPVDSTYIVSR
HGAFAASADVNSALITGTYRFVSPFGRTIFARNDSTWGAKPMDKDFERIA
NANRRPALADVDVIDKSPDRLKTSVNVFAVQGRALIIQSPFRGPVPC
IDNKKPLAENPILDVVRKALSLAIRPADERMQGTVAHOMPMANRPNVPC
IKNIPDVKAQDLAQGFPEGQTLTHVPEGRIPOAEVQMAVAQVTRIGYKQVL
EVLPAVAVVAGKAKNELAISVLAQNGTEAAVALTNIITVDSKGSQSWGHYSN
PLVDKALADSTAEFDEAKRRKILBDSVQVSDVGIIPLVHQNIMDKGLKVEPLV
SDRTAIVTVEQRP"
7255. .7776
/locus_tag="PSPTO2815"
/locus_tag="PSPTO2815"
7255. .7776
/locus_tag="PSPTO2815"
/note="similar to GP:15162222; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA056314.1"
/db_xref="GI:28853246"
/translation="MTTESNSPPDLDLHLGIEPGNPLHTVHARDKVASATQSQAL
FEDPALADNLPLESERLWAVYATQISAOPTLNAVYIQQREGVADSVADYDANID
LLEDPRILALIFRRTILSPHVGDSALQALQGLSTAEIVLAQLAFLASYQVRL
AAGLTAIQSAGAA"
7776. .8366
/locus_tag="PSPTO2816"
/locus_tag="PSPTO2816"
7776. .8366
/locus_tag="PSPTO2816"
/note="similar to GP:15162222; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="alkylhydroperoxidase AhpD domain protein"
/protein_id="AA056315.1"
/db_xref="GI:28853247"
/translation="MSLEIEIGFTNQQVIGKAMLPYDNLNATYAEQTAVEESHPQA
RTSDVYTLAHPDILRQSRQAFNAIMVAPGSLAEERLASTVVSINRCVYCASVH
AQRPEOLAKRNDVIAQVADPACTAGTARAKAIVQFAIDLTLPEASLNAHIIALREQ
GLDDAQVLDLIIAIAIFAWANMLMNLGSPVPEAE"
8429. .8614
/locus_tag="PSPTO2817"
/locus_tag="PSPTO2817"
8429. .8614

/locus tag="PSP02817"
 /note="Similar to GP:15158870, and GP:15980012; identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="conserved domain protein"
 /protein_id="AA056316.1"
 /db_xref="GI:28853248"
 /translation="MAECHPSLPMVCARVSGATVTEPRKARTMHTPLIGQLIVNAGAGVSDAETASGAPQ"
 complement (8591..9982)
 /locus_tag="PSP02818"
 complement (8591..9982)
 /locus_tag="PSP02818"
 /note="Identified by match to Pfam protein family HMM PF01569"
 /codon_start=1
 /transl_table=11
 /product="fatty acid desaturase"
 /protein_id="AA056317.1"
 /db_xref="GI:28853249"
 /translation="MYLVLTGLVLTVTHVPMGSVRLVQPSVDAAHVPLLPFLPYLSTLVMPVIVVMGROSMLPTFPAGLAAGLCVCHLFWPTMLRBTGAAMLDWLYQDAPLAASPSGHALPAVASVAMAGRLQKAVFWASVLMLTMTTGOHFTDMA YGAIGVACGAATLIPKRCADVLRITLSALLEWLCITLRVLYLADMRPYLLMLI VVAROHALFVLYNDATHNYLTERRSTNDPLINLAIGVGLVPIEFYRPLLDHNOHAG TEODERRRFLYTRQPMRFQPLSKLRLQLDGLMWNTRIMATKAAAGAPKPTR PLISALVMDLITVACLWQCSIREVGMLAMWFLPVATGTLQKRSIAESHGSGEV TRPMQMTYAMRWGMLGRFPIWYHNLQHRHASIPWHALPSAVGDEKMLPSRA LPLMWSGLRRKT"
 10166..10954
 /locus_tag="PSP02819"

gene

Alignment Scores:
 Pred. No.: 2, 09e-164 Length: 310029
 Score: 2474.00 Matches: 460
 Percent Similarity: 87.48% Conservative: 50
 Best Local Similarity: 78.90% Mismatches: 73
 Query Match: 80.27% Indels: 0
 DB: 1 Gaps: 0

US-09-750-986D-30 (1-589) x AE016866 (1-310029)

QY 1 MetArgSerLeuGluAlaLeuLeuProRheProGlyValGlyIleLeuGluArgLeuGluHis 20
 DB 172246 ATGCGAGCCCTGGACCCGTGGCGCGCTGTCGCGATGCGTGTGCGACCGCGCTGGCGAC 172187
 QY 21 TrpAlaLeuThrArgProGluGluGlnThrCysValAlaAlaArgAlaAlaAengIyGluTrp 40
 DB 172186 TGGGCCAGCGCTGGCGCAACAACCTTATATGCGCCGCTGACAGCCCTGGTGGCTGG 172127
 QY 41 ArgArgIleSerTyraAlaGluMetPheHisAsnValArgAlaIleAlaGlnSerLeuLeu 60
 DB 172126 CGCAAGGTCAAGTACCGCCGACATGCTCAACGATGCGCCCATGCGCCAGACCTGCTG 172067
 QY 61 ProTyGlyLeuSerIleGluArgProLeuLeuIleAlaSerGlyAsnArgLeuGluHis 80
 DB 172066 GCTTACGGGCTTTCGCGAGGCGGCCACTGCGCTGTCGGGTAAACGACATCGACAT 172007
 QY 81 LeuGlnLeuAlaPheGlyAlaMetTyraAlaGlyIleProTyCysProValSerProAla 100
 DB 172006 TTGCAATGCTTGGCGCGCTATGTACGCGGATTCCTTATTCGCCGATGTCAACGGCT 171947
 QY 101 TyrSerLeuLeuSerGlnAspLeuAlaLeuLeuArgHisIleValGlyLeuLeuGlnPro 120
 DB 171946 TACTCGGTCTGTGCGAGGACTTTCGCAAAATTCGGCGATGTCTGTGAGGTGTTACAGCCA 171887
 QY 121 GlyLeuValPheAlaAlaAspAlaAlaProPheGlnArgAlaIleGluThrIleLeuPro 140
 DB 171886 GGCCTGCTGTTTTCGCGAGTGGCGCGCTTCACGCGCCACATCGATGCTGCATACCG 171827
 QY 141 AspAspValProAlaIlePheThrArgGlyGlyLeuAlaGlyArgArgThrValSerPhe 160
 DB 171827

DB 171826 GTCACACACCGGTCACTACGTACCGGACAGTTGAATGCGCGCTCATTCGAGCTTT 171767
 QY 161 AspSerLeuLeuGluGlnProGlyGlyIleGluAlaAspAsnAlaPheAlaAlaThrGly 180
 DB 171766 GCCAGCTGCTTAAACAACCGGGCGGACGGAAGCCGACCTGCTCATGCGCAGCGGG 171707
 QY 181 ProAspThrIleAlaLysPheLeuPheThrSerGlySerThrLysLeuProLysAlaVal 200
 DB 171706 CCGGACAGCTTCCCAAGTTCTTTTCACCTCGGGCTCAACCAACTGCCCAAGACATG 171647
 QY 201 ProThrThrGlnArgMetLeuCysAlaAsnGlnIleMetLeuLeuGlnThrPheProVal 220
 DB 171646 GTTACCACTACGCGCATGCTGTGCGCAATCAAGATGCTGTGCAAACTTCCCGGTG 171587
 QY 221 PheGlyGlyGluProProValLeuValAspTrpLeuProThrAsnHisIleThrPheGly 240
 DB 171586 TTGCGGAAACGCGCGCGGTGTGAGATGAGTCCGCGGAATACCTTCGCGGCG 171527
 QY 241 SerHisAsnIleGlyIleValLeuTyraAsnGlyGlyThrTyTrpLeuAspArgLys 260
 DB 171526 AGTCATTAACCTCGGTATCGTCTTACAGCGCGGACGCTTCTTCCGCAATGCGCAA 171467
 QY 261 ProThrAlaGlnGlyPheAlaGluThrLeuArgAsnLeuSerGluIleSerProThrAla 280
 DB 171466 CCGACTGTTACAGGCTTTGGCGAAACCTTGCACACTCAAGAAATCTCGCGACAGCC 171407
 QY 281 TyrLeuThrValProLysGlyTrpGlyGluLeuValGlyAlaLeuGluValAspSerThr 300
 DB 171406 TACCTGACCTGCGCAAGCGCTGGGAAGACTGTGTAAACCGGTGGAGCGAGCGGAG 171347
 QY 301 LeuArgGlyArgPhePheAlaArgMetLysLeuPhePheAlaAlaAlaGlyLeuSer 320
 DB 171346 TTGCGTAGCGATTCCTTCGCGCATGAACTTCTTCCCGCGACGCGGTTCTCG 171287
 QY 321 GlnGlyIleTrpAspArgLeuAspArgValAlaGlnGlnHisCysGlyGluArgIleArg 340
 DB 171286 CAATGATATCGGACCGCGCTGCAGCGGGTGGCGACAGCATGTGGCGAGCCATTCGC 171227
 QY 341 MetMetAlaGlyLeuGluGlyMetThrGluAlaProSerCysThrPheThrThrGlyPro 360
 DB 171226 ATGATGCGCGGTGGGACAGACCGAGCGCGACCGCTTCGACCTTATCACCGGGCG 171167
 QY 361 LeuSerMetAlaGlyTyrrIleGlyLeuProAlaProGlyCysGluValLysLeuValPro 380
 DB 171166 CTGTCAATGCGCGCTTATCATCGCTTGCAGCGCGCGCGGAGCGAGTGTCTGTGTCGG 171107
 QY 381 ValAspGlyLeuLeuGluGlyArgPheHisGlyProHisValMetSerGlyTyrrTrpArg 400
 DB 171106 GTTGATGGCAAGCTGGAAAGGCGCTTTCGCGCGCGCACATCATGCTGACTGGCGC 171047
 QY 401 AlaProGluGlnAsnAlaGlnAlaPheAspGluGluGlyTyrrTyrcysSerGlyAspAla 420
 DB 171046 TCGCCGACGAAACCGCTGAGGTTCGAGAAACGCGCTTATGCTTCGCGTGAACGCG 170987
 QY 421 IleLeuLeuAlaAspProAlaAspProGlnLysGlyLeuMetPheAspArgIyArgIleAla 440
 DB 170986 ATCAAGCTGACAACTTTCGCGCGCGCGACCTGAGTGTTCATGAGGCGCTTGGCC 170927
 QY 441 GluAspPheLysLeuSerSerGlyValPheValSerValGlyProLeuArgThrArgAla 460
 DB 170926 GAGGATTTCAAACTGTCGTCGCGGTGTGTCAGCGTGGGCAATCGCAATCGGCGC 170867
 QY 461 ValLeuGluGlyGlySerTyrrValAlleuAspValValAlaAlaProAspArgGluCys 480
 DB 170866 GTGCTGAAAGCTCGCGCGATGTTCAGGAAGTGTATCAACCGACCGGATCGCGAGTGC 170807
 QY 481 LeuGlyLeuLeuValPheProArgLeuLeuAspCysArgAlaIleLeuSerGlyLeuGlyLys 500
 DB 170806 CTGGGGCGCTGTTCTTCTCCCTGCGACAGATGCGCAACATTCGCTGGCGCTGGGTCG 170747
 QY 501 GluAlaSerAspAlaGluValLeuAlaSerGluProValArgAlaTrpPheAlaAspTrp 520
 DB 170746 CAGCGAGCGATGTCAAGTATGGCCAGACCGCGGTACGCGAGTGTTCGCGACTGG 170687

QY	521	LeuysaAgileusAsparaglinAthrgrgysanaAseArgilemetrPvA1gIyleu	540
Db	170686	CTGGCACCTTTGAATGAACGGGGCCAAATGCAACGCCAGCTGCTGAAGTGGATCGCGCTG	1706227
QY	541	LeuAspThrProProSerIleAspIlysgIyguValThAspIlysgIyserIleAengIn	560
Db	170626	CTGATATGAACCGCCTCAATGCATGCAGCGCGCAAAATACACTGACAAAGGGCTTATCAACCG	1705667
QY	561	ArgAlaValIleuGlnTTPArgSerAlaIysValAspAlaLeuTYrArgIyguAepGln	580
Db	170566	CGTGGGTGTTTGCACATGGCGCCGACGAAAGTCGAAGCCCTGATGCATCAGATGCC	1705070
QY	581	SermIleu	583
Db	170506	TCACGGCTC	170498
RESULT 5	PFL536325	9477 bp	DNA linear BCT 09-JAN-2004
LOCUS	DEFINITION	Pseudomonas fluorescens febr gene, ech gene, vdh gene, fcs gene, aat gene and fcd gene.	
ACCESSION	AJ536325		
VERSION	AJ536325.1	GI:40788116	
KEYWORDS	aat gene; beta-ketochiolase; ech gene; fcd gene; fcs gene; febr gene; feruloyl-COA dehydrogenase; feruloyl-COA hydratase; feruloyl-COA synthetase; transcriptional regulator; vanillin dehydrogenase; vdh gene.		
SOURCE	Pseudomonas fluorescens		
ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.		
REFERENCE	1	Barghini, P., Ciovolani, C., Picca, A. G., Schieser, A. and Ruzzi, M.	
AUTHORS	Cloning and Characterization of Ferulate Catabolism Genes from		
TITLE	Pseudomonas fluorescens Bfl3		
JOURNAL	Unpublished		
REFERENCE	2	(bases 1 to 9477)	
AUTHORS	Ruzzi, M.		
TITLE	Direct Submission		
JOURNAL	Submitted (09-JAN-2003) Ruzzi M., Agrobiology and Agrochemistry, DABAC - University of Tuscia, via C. de Vellis, snc, Viterbo - 01100, ITALY		
FEATURES	Location/Qualifiers		
source	1..9477		
	/organism="Pseudomonas fluorescens"		
	/mol_type="genomic DNA"		
	/strain="Bfl3"		
	/isolation_source="soil"		
	/db_xref="taxon:294"		
	/environmental_sample		
	/country="Italy:Lazio"		
	complement(1297..1785)		
	/gene="Febr"		
	complement(1297..1785)		
	/gene="febr"		
	/codon_start=1		
	/evidence=experimental		
	/transl_table=1		
	/product="transcriptional regulator"		
	/protein_id="CAD60265.1"		
	/db_xref="GI:40788117"		
	/translation="MAKPAPLADPREAPPANNAVOAPLSDALDELIGYAMRRQALKF		
	ONLIGOLSAHDRPAQFSALATITDSNPGMOADLAPALITPPVAVPLINKLESSALA		
	VPRVCKPDKRSIGTFLSKTGTLELKEIKOIAAGSDVATSAESREBELRLIKVY		
	OD"		
	1994..2824		
	/gene="ech"		
	1994..2824		
	/gene="ech"		
	/codon_start=1		
	/evidence=experimental		
	/transl_table=1		

[illegible]

/product_id="fernyoy1-coa hydratase"
 /protein_id="CAD60266.1"
 /db_xref="GI:40788118"
 /translation="MNSYSEKRWTTVKEVIEDGIAMVILNPKRNAMSPTLNREKIDV
 LETLEGPAGVILVIGAGAGYAGAMDLKXEYFPEYDAGEPIIQEKIRBASQWOKLIL
 RMYAKPTVGVKSGESFPVACDLACADAEFGISBINMI PGNIYSKAMD
 TVGRHOSLYVIMSGKFGGOKAEMGLVNDSPVLRIRVTELRNLKRNKPVLRK
 AKGFKRCRELITMEQNEEDLYALUDSRLLDTPGREGQMKGFLDDKSIKQGLAYKKA
 "

2921. 4369
 /gene="ydh"
 2921. 4369
 /gene="ydh"
 /codon_start=1
 /evidence="experimental"
 /transl_table=11
 /product="vanillin dehydrogenase"
 /protein_id="CAD60267.1"
 /db_xref="GI:40788119"
 /translation="MLDVPILITGGGRPARDOGRTLERNPVTGNAVSVAAATLDDAD
 AAVAAQAPFAMPAWALPNRRSRILKAAEQIQAQGEFTIAAGSTGMAMWYGNVNR
 LAANMLREASMTQTQINGEYIPSDYSGSPAMLRQPCGVVIGIAPMNPVILATRAIL
 PMVRNVTGTVTKSGEISPAVHRILIGVLLDAGIGGVVIVISNAVDAAALYERILANM
 PLACVAGTSTVTHGRIVGELARHLKPLILLEGKGAPLVVDADLDAAVAALAFGA
 RVGQOICSMREPLINDAKVDAFVAKAAKVAVALRAGDPSAADSGSLVDVSAGRF
 IKALIDILSKGARILVVGQDGSITIOPTLIVGTSENRILYRESFGVYVALLDPA
 EELLRLANDSEFGISAIIFSRIDGRLLALAQVSESCIHINGPVHDEAOMPFGVGS
 SGVSGFGKASIEHFTQLRWVTLQHPRHYPPI"
 4628. 6397
 /gene="Fcs"
 4628. 6397
 /gene="Fcs"
 /codon_start=1
 /transl_table=11
 /product="fernyoy1-coa synthetase"
 /protein_id="CAD60268.1"
 /db_xref="GI:40788120"
 /translation="MNSLRKFLALPRLILDLRYHAKVRPEQPTIAAEBAQDRLVYS
 YGQMLDSVRAIAQGLRLRYGLSADKPLALLSGNDIEHLQALGAMTAGIIPYCPSPAYS
 LLSQDPKLRHVCDLQPGVLFVSDASAVORAIDNVLPEPTPLISRVQVGRQAS
 ASLAEPCGAEMDAFAATGPDISIAKEFLTSGSTKLPKAVITTOQLMANQOMLLQTL
 PVFGEPEPVLVDMLPMNHTFGSGSHVGVILVNGGPFYLDDEKPTAQGFALTRNILE
 SPYATVLPKMEIIVVVALEEQDALEDECFPKMSLFFPRAAGLSQSWDRLDVALEGE
 CGERTRMWAGIGMETEAPSCFTPTGPLSNAGTIGLPAPCEVRLPVVDKLEGGFRGC
 HIMGEYMRALQDALEVDAQVFCSGDALADADADPQGLMFDRILADEKFLSGVGF
 VSVGPELRNALTECAPVQDQVLAADRECLALVFPRLYECRRILRAGABASEAOVY
 ASAPYRWGFGMLQCLNRRASGNASLEWIALQVEBASIDRGEITDKGSIQORAVLQW
 RASQYRWGFGMLRRESILRAEPRK"
 6394. 7641
 /gene="aat"
 6394. 7641
 /gene="aat"
 /codon_start=1
 /transl_table=11
 /product="putative beta-ketochiolase"
 /protein_id="CAD60269.1"
 /db_xref="GI:40788121"
 /translation="MMSGAGYSGFADVAALIEAVRTPWVDLGGDLAEISPIDLGIKVGR
 EVLRAQGEPOAVDYSVLAGMSASPDVALPHHIGLYSGVGAOAPALGAVORICATGEG
 ELIQAQAEQRAEYQVQLCAYSASMRNPILAYTHRRGQFGRGLAAQOPDLMALYOL
 ALGDMITATDANLRRKRGTLREAVDYGACDSHQRLAAKQAMPAALIVANSQRFEL
 DGYQARGRITPRGEVRSQDSHPERSERDILARLTIEGGVQTQAGNSCAVDAALAA
 LVGASACSDPDLALASAVVNAVEFMGIGAPARILLOKRPBAFGQIRSAFPI
 NKAQASGVLAAPFLGIDGGRNLVAVGGSIALGHPILAAAGSLRVLTLARQLEHNLRYG
 IAAQAGVGQGGOMLLIENPMWR"
 7701. 9455
 /gene="Fcd"
 7701. 9455
 /gene="Fcd"
 /codon_start=1
 /transl_table=11
 /product="putative fernyoy1-coa dehydrogenase"
 /protein_id="CAD60270.1"

```

/db_xref="GI:40788122"
/translation="MMSYQAPLADMOFVLEHMLAPDAPRSPAFALDLPALQVLE
QARPSQVLAIPRIANASGDRCRMADQSVSTPSPGADATRAYIEGGPALACBAALGG
OGIPQLDLALQEMLYASNTAMWMTQVLAIGAVLCUKTRHAPMLQERYPAIIISGLSL
PMTCLTEPLASLVGLRCAEPQADQSVYLSKLFISGGEHLDNDNLIALVLAIP
DAPEGSGISLFLVPRLEQGDNGVRCNCHIEHKMFKISGATGSLVFDAQGVLEBA
NRGLAMFWMNNSARLHVLOGIYHVAAMQNALADAGEOQKARPLRPAQAQAD
PIRYHPAMRVVLELRATSRGLAIGVMAHLDOAEHNEPRLARRQADLQDLPI
ITAFTEQGRFASNLQVTFGYVAEPFIEOTLDRSIAIYESNEIQANDLLIR
KVLGDGQGFALIEQURDEAEGAAVAGGAGPLADIAELQGLVGEVQWALDE
AEYPRAADFLALCGLSLALAMARAVSRNLPGDPLRGAKLTAGFFSYLLPQ
VEQRMTAVBSQAALAFI"

```

ORIGIN

```

Alignment Scores:
Pred. No.: 1,35e-164 Length: 9477
Score: 2449.00 Matches: 456
Percent Similarity: 87.03% Conservative: 54
Best Local Similarity: 77.82% Mismatches: 76
Query Match: 79.46% Indels: 0
DB: 1 Gaps: 0

```

US-09-750-986D-30 (1-589) x PFL536325 (1-9477)

```

QY 1 MetArgSerLeuGluAlaLeuLeuProPheProGlyValThrLeuGluAlaGluGluHis 20
DB 4628 ATGCGCTCGGTGAAGCGCGCTTGGCGGCTTGCCTTCATCGCTGCTGTCAC 4687
QY 21 TTPAlaLeuThrArgProGluGlnThrCysValAlaAlaArgAlaAlaAnglyGluTrp 40
DB 4688 TGGCGCGCGCGCGCGCGCGAGACTTCATTGCCGCGCGAGCGCGGTGGCGACTGG 4747
QY 41 ArgArgIleSerTyrrAlaGluMetPheHisAsnValArgAlaIleAlaGlnSerLeuLeu 60
DB 4748 CGCGTGTGCGCTATGCGGAGATGCTCGACAGCGGTGGCGCGCATTCGCCCGCGCGCTT 4807
QY 61 ProTyrrGlyLeuSerAlaGluArgProLeuLeuIleValSerGlyAsnAspLeuGluHis 80
DB 4808 CGTTACGGCGCTGCGCGCGACAGCGCGCTTGGCGCTGCTTCGGGCAATGACATCGAACAC 4867
QY 81 LeuGlnLeuAlaPheGlyAlaMetTyrrAlaGlyIleProTyrrCysProValSerProAla 100
DB 4868 CTGCAACTGGCGCTGGCGCGGATGTACGCGGGGATTCCTATTGCCGCGGTGCGCGGCC 4927
QY 101 TyrrSerLeuLeuSerGlnAspLeuAlaValLeuArgHisIleValAlaGlyLeuLeuGlnPro 120
DB 4928 TATTTCCTGCTGCTCCAGGACTTTGGCCAGCTTGGCGGCATGTCTGCGACCTGTGGCAGCCG 4987
QY 121 GlyLeuValPheAlaAlaAspAlaAlaProPheGlnArgAlaIleGluThrIleLeuPro 140
DB 4988 GGGCTGGTGTCTCTCAAGCGCGCGCGCTTACAGCGGGCATTCGACGCGGTGCTGCCG 5047
QY 141 AspAspValProAlaIlePheThrArgGlyGlyLeuAlaGlyArgArgThrValSerPhe 160
DB 5048 CCGAGACCGCGCGCTGATCAGCGTGGCGGAGGTGCGCGGGCGCGCGACGAGCTTC 5107
QY 161 AspSerLeuLeuGluGlnProGlyGlyIleGluAlaAspAsnAlaPheAlaAlaThrGly 180
DB 5108 GCCAGCTGCTGGCGCGCGCGGTGGCGCGAGCGCGCATCGCGCTTTGGCGCGACCGGC 5167
QY 181 ProArgThrIleAlaValPheLeuPheThrSerGlySerThrIleValLeuProValAlaVal 200
DB 5168 CTTGACAGCTGCCCAAGTTTCTTCACTCGGGCTCGACCAAGCTGCCCAAGCGCGGTG 5227
QY 201 ProThrThrGlnArgMetLeuCysValAsnGlnGlnMetLeuLeuGlnThrPheProVal 220
DB 5228 ATCACCACCCAGGAGCATGCTCGCGCAACAGACAGATGCTGCTGCGACCTTCCGGTG 5287
QY 221 PheGlyGlyGluTrpProProValAlaAspTrpLeuProThrPheAsnIleThrPheGlyGly 240
DB 5288 TTTCGGTGAAGAGCGCGCGGTGCTGAGATGCTGCGGTGGAACCAACATTCGCGGCG 5347
QY 241 SerHisAsnIleGlyIleValLeuTyrrAsnGlyGlyThrTyrrIleuAspArgGlyLys 260

```

```

DB 5348 AGCCATPAGCTGGCATTCGCTCTACACGCGGGATCTTTTACCTGACGAGGGCAG 5407
QY 261 ProThrAlaGlnGlyPheAlaGluThrLeuArgAsnLeuSerGluIleSerProThrAla 280
DB 5408 CCCACGCGCCCAAGCTTCCTGTAACCTCGCGCACTTCAGGAGATTTCCCCACCGCC 5467
QY 281 TyrrLeuThrValProIleGlyTrpGlyGluIleValAlaIleGluIleArgAspSerThr 300
DB 5468 TACCTGACCGCTGCCCAAGGCTGGAGAACTGTCAACCGCTTGAGAGAGCGCGAG 5527
QY 301 LeuArgGluArgPhePheAlaArgMetLysLeuPhePhePheAlaAlaAlaGlyLeuSer 320
DB 5528 CTGCGGAGAGTGTTTTTCAGGCGCATGAGCTGTCTTTCTTGGCCCGCGGCTTTCG 5587
QY 321 GlnGlyIleTrpAspArgLeuAspArgValAlaGluGlnHisCysGlyGlyAlaGlyIleArg 340
DB 5588 CAAGGTGCTGGGATGCGCTGGACCGCGGTGCGCGACAGCATTTGCGCGCAACCATCCGC 5647
QY 341 MetMetAlaGlyLeuGlyMetThrGluTrpAlaProSerCysThrPheThrThrGlyPro 360
DB 5648 ATGATGCGCGGCTCGCATGACCGAGCGCGCGCTTGTGACCTTCACACCGAGCCG 5707
QY 361 LeuSerMetAlaGlyTyrrIleGlyLeuProAlaProGlyCysGluValLysLeuValPro 380
DB 5708 CTGTGATGCGCGGTTCATCATCGGCTGCGCGCGCGGGTTGGAGATGGCGCTGTGCGG 5767
QY 381 ValAspGlyLysLeuGluGlyArgPheHisGlyProHisValMetSerGlyTyrrTrpArg 400
DB 5768 GTGATGCGCAACTCCAAAGGCGTTTTCGGCGCGCGCACATCATGCGCGCTCTGGCGC 5827
QY 401 AlaProGluGlnAsnAlaGlnAlaPheAspGluGluGlyTyrrTyrrCysSerGlyAspAla 420
DB 5828 GCGCGGAGAGACCGCGCGAGGTTCGAGCGCCAGGCTTCTACTGCTTCGGGGATGCG 5887
QY 421 IleValLeuAlaAspProAlaAspProGlnLysGlyLeuMetPheAspGlyArgIleAla 440
DB 5888 CTCAGCTGCGCGACCGCGCGCGCGCAACGCGGCTGATGTTCCAGCGCGCGCATTCGCC 5947
QY 441 GluAspPheLysLeuSerSerGlyValPheValSerValGlyProLeuArgThrArgAla 460
DB 5948 GAGGACTTCAGATGTCGTCGCGGGGTTCCTCAGCGCTCGGACCGCTGCGCAATCCGCG 6007
QY 461 ValLeuGluGlyLysSerTyrrValLeuAspValValAlaAlaProAspArgGluCys 480
DB 6008 GTCTCGAAGCGCGCGCTATGTTCAGGACTGTGTGTGCGCGCGCCCGCGACGCGAGTGC 6067
QY 481 LeuGlyLeuLeuValPheProArgLeuLeuAspCysArgAlaLeuSerGlyLeuGlyLys 500
DB 6068 CTGGGGCGCTGTGTTCCCGCGCTGTATGATGCGCGGCGCTGGCGGGGCTGGGCGCC 6127
QY 501 GluAlaSerAspAlaGluValLeuAlaSerGluProValArgAlaTrpPheAlaAspTrp 520
DB 6128 GAGGCGACGAGCGCGCAGGTGCTGGCGAGCGCGCGGTGGCGCAGTGTTCGCGACTGG 6187
QY 521 LeuLysArgLeuAsnArgGluAlaThrGlyAsnAlaSerArgIleMetTrpValGlyLeu 540
DB 6188 TTTCAGTGGCTGATGTGAGGCTCAGCGGCAATGCGAGCGCTGAGTGTGAGCGCTG 6247
QY 541 LeuAspThrProProSerIleAspLysGlyGluValThrAspLysGlySerIleAsnGln 560
DB 6248 CAGGTGAGCGCGCGCTCATTCGACCGTGGGGAGATCACCGACAAAGGCTGATCATTCAG 6307
QY 561 ArgAlaValLeuGlnTrpArgSerAlaLysValAspAlaLeuTyrrArgGlyLysAspGln 580
DB 6308 CGCGCGGTGCTGAGTGGCGCGCGCGAGTGCAGAGTGCAGAACTGTATCGCGCGCGAGCCG 6367
QY 581 SerMetLeuArgAspGlu 586
DB 6368 TCGATCTCGCGCGCGAG 6385

```

RESULT 6
AB016786/c

LOCUS AE016786 301214 bp DNA linear BCT 12-DBC-2002
DEFINITION Pseudomonas putida KT2440 section 13 of 21 of the complete genome.
ACCESSION AE016786 AE015451
VERSION AE016786.1 GI:26557030
KEYWORDS
SOURCE Pseudomonas putida KT2440
ORGANISM Pseudomonas putida KT2440
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE
1 (bases 1 to 301214)
Nelson,K., Paulsen,I., Weinel,C., Dodson,R., Hilbert,H., Fouts,D.,
Gill,S., Pop,M., Martins Dos Santos,V., Holmes,M., Brinkac,L.,
Beanan,M., Deboy,R., Daugherty,S., Kolonay,J., Madupu,R.,
Nelson,M., White,O., Peterson,J., Khouri,H., Hance,I., Lee,P.,
Holzapfel,E., Scanlan,D., Tran,K., Moazzez,A., Utterback,T.,
Rizzo,M., Lee,K., Kosack,D., Meestl,D., Medler,H., Lauber,J.,
Hohnselt,U., Straczek,M., Heim,S., Kiewitz,C., Eissen,J., Timmis,K.,
Duesterhoft,A., Tummeler,B. and Frazer,C.
Complete genome sequence and comparative analysis of the
metabolically versatile Pseudomonas putida KT2440
Environ. Microbiol. 4 (12), 799-808 (2002)
2 (bases 1 to 301214)
Nelson,K., Paulsen,I., Weinel,C., Dodson,R., Hilbert,H., Fouts,D.,
Gill,S., Pop,M., Martins Dos Santos,V., Holmes,M., Brinkac,L.,
Beanan,M., Deboy,R., Daugherty,S., Kolonay,J., Madupu,R.,
Nelson,M., White,O., Peterson,J., Khouri,H., Hance,I., Lee,P.,
Holzapfel,E., Scanlan,D., Tran,K., Moazzez,A., Utterback,T.,
Rizzo,M., Lee,K., Kosack,D., Meestl,D., Medler,H., Lauber,J.,
Hohnselt,U., Straczek,M., Heim,S., Kiewitz,C., Eissen,J., Timmis,K.,
Duesterhoft,A., Tummeler,B. and Frazer,C.
Direct Submission
TITLE
Submitted (05-NOV-2002) The Institute for Genomic Research, 9712
Medical Center Dr. Rockville, MD 20850, USA
FEATURES
source
1..301214
/organism="Pseudomonas putida KT2440"
/mol_type="genomic DNA"
/strain="KT2440"
/db_xref="taxon:160488"
complement(69..1007)
/locus_tag="PP3196"
/note="similar to GB:X79440, SP:Q16798, and PID:535012;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="modulation protein dl"
/protein_id="AAN68804.1"
/db_xref="GI:24984828"
/translation="MRFHLDLNLVALDILLEONITRAERLHMTQSATSGVLGRLL
RPFPEELIVGVGRKQPTTPVALTELAKPVREVLITRASITAKRPVDPAPASHPFLV
ADYLISVFAOVYOKIHOEAQITPEMIGPQSSSMLMRGEVDMYIPERYLIEGH
PAQLFEEEDVACVMKRDNPVEVSHLTLEQYMOGHI SVFGKRRHLISIDMFTNQGFG
NRRLEVTINDENFTLLPQLLIVGTAVATMRRLRLVAGYPLRLPPVVKI PVMREBML
WHRSMGDGPMHRLRERISDVIOCASTQDALRATA"
1284..2225
/locus_tag="PP3197"
1284..2225
/locus_tag="PP3197"
/note="similar to GB:X54232, SP:P35052, and PID:31847;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="glyoxalase family protein"
/protein_id="AAN68805.1"
/db_xref="GI:24984829"
/translation="NNILGLDLVFGVDDIOACADCLRDYGLDAVQDLDDGGRFEALD
GSIITIRGDDPLPPALPPTPSIRFTYVGVASSTDLAIADLIGDRPVRDASGAL
HHPDILGFALFQVTCRRPBDAPDITNAPGAPORPTNOLGIGAPMRPALPTLSLVV
YVPPDAKGEATYRDLGFTVTTDSFIVGPPFKMTAMDHHCIFMLQTPPHNQCCHP
TTHMSGTIVLLAGTRFQOQWTSFVGPGRHLGSGWFWYFNSPLGCHLEYDADMDKH
DAAWAAQAPMSADNSQLFLFASREKMASSGPAAKOG"

gene
2238..2576
/locus_tag="PP3198"
2238..2576
/locus_tag="PP3198"
/note="identified by match to PRAM protein family HMM
PF00355"
/codon_start=1
/transl_table=11
/product="ferredoxin, 2Fe-2S"
/protein_id="AAN68806.1"
/db_xref="GI:24984830"
/translation="WELCLDELGGGQARGFPOGRGADSLFALRYLGVRYYRLCP
HLMVPEYRKRNFLADQWIVCAHGAFTPEDTGVHGPGRGAELELDHKEAGWG
LLPLPEQDR"
2716..3840
/locus_tag="PP3199"
2716..3840
/locus_tag="PP3199"
/note="similar to SP:P23262; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="monooxygenase, putative"
/protein_id="AAN68807.1"
/db_xref="GI:24984831"
/translation="NNTVNVYLIVGGIGGLCAIATLRKGIADVLDVELKTEMTYGV
GIIOOSNVYREHRLGLEAYLDAVAAYFEDVAYITTEGOQLARIQORLAGQYANV
GISRAHLKVISSETAIEIGTOYRLGTSVESPEQDSADVDFPTDSQGRYGLMWGADG
LFSKRSVLFCGRYOPRFTGQSVWYRNPPRAGADHILANVQSPGNCAGIYPLADLMY
LFTTSEPGNPMFEPALADGLRKLAFGGILGRLREQIVNAGVYKPELAAYTDE
PWYCGRVLLIGDAALHTPHLGAGAGMAIEDVLALELTAGSGLDAOLQRMGRKFE
RCKFSEMSLLAGVKEIQRPSPFRIGLVKMLVETADPI"
3885..5597
/locus_tag="PP3200"
3885..5597
/locus_tag="PP3200"
/note="similar to GP:3184126, and GP:3184126; identified
by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAN68808.1"
/db_xref="GI:24984832"
/translation="MPTPHNKKKEETPMLTRKGAFAAPLPALKEPLSLVALAASQA
QAFEDTGNPLDRVMDNTLKYSAMARTODASSKTEGDTALNDGDBNPFKHLISN
RDLISLSDITNNVNGARVSGAAYDDVYNOGTNNDDPGRANSYSDVDAFPDDRLK
HGRKEILDARIFGKTEIGDMPLSGRLQVAMQGESLFEKNGGIAAGWAPDVYKGL
SVPTQFKELIRPVQIISGQLQITPDVSVGAYIQRBEANRLPAAGSYFTSTDPFGDS
AERMPVGAAPFPSPGQPLAFPHGNDKQAKDSGGGQVRRRTETVDMGLYAIRFHDKSP
QLNVRPNFANDLPVSGKAGEFYWVYPEGIEAFGASFTVGNFNLAGELISTMNQPLA
STNGHALVGEAIDDDNPVYATGRTLHANYSLASLEBNFYAESSFAGEIAMNRVM
SVSKNADVPADADYSIRLVPEPMYRQILPGDLISVIGASVTNGSMALMGCFG
TDHGQDNLNIGLRNVLNTWSLGLYTHHYGPENTPLDADNHTTFQGS�KDRFLAFTV
SRTP"
7034..8095
/locus_tag="PP3201"
7034..8095
/locus_tag="PP3201"
/note="identified by match to PRAM protein family HMM
PF02012"
/codon_start=1
/transl_table=11
/product="BNR domain protein"
/protein_id="AAN68809.1"
/db_xref="GI:24984833"
/translation="NNKMLYGMWLLAACGCLPPAFAGDSYVLAQPALHQPALRAVLO
DVAARDARLVAVAGERGVLISDDNATWQQAIAVAVGSLVAVQFIIDANGVAVGAG
VYLHODDGEHWSVOLDGKRAALAEVQAAEAADRRRLAAQRLADGADKPLALNF
TDARNGLVGAAGVAMRTDDGGRSMQSEIGRLPNRGLHLYALAQGVYTAGCGL
ILRSDDGHEHYALGPYDGSFEATVLPDGLLGGRLGSLFASDDGDSYAGOLANP
TPASINGCARVAGQOLLANQGMULARSGIGAFAPAAQPAVAVSDCLPLTAATTADSGIVA
VMAQAKRRLESLSITTD"
8108..10588


```

CDS
    /locus_tag="PP3202"
    8108..110588
    /locus_tag="PP3202"
    /note="similar to GP:9949187, identified by sequence
    similarity; putative"
    /codon_start=1
    /transl_table=11
    /product="transporter, putative"
    /protein_id="AA068810.1"
    /db_xref="GI:24984834"
    /translaction="MNSAPSTGTPDPDSAGSLERAFNHRVLTLCATLTL
    GQASRLTLNASEPKMIRDPPIHNYEROLAGNVRIVANPRISIDKYL
    HSIQOLNDAYLTVGVBAKMSLMTSTWTGTEGGVGPVPGVGAASLEA
    LKRNVERSNETIGQVAFDROSSIVVPLLEKTPDGOALDYTAHELETRERPOAG
    VEHTGPAKVGDMDGLKQITLFPAAIAITTAATVATYTRCVSTALVYVCSLAV
    IWQGLPLGLVLDPSVLPVFAIGMSHAQKXNGIMODIGRMRLVARLTP
    RRLFLAGLTALCDVAFVLMITQIQLVDLVIASLGAVALFNLILPLLSV
    GSARARSRLRAEASAGAGAVRFDLFTRRMAACIAVALMAAGYAVSL
    HIKVGLDGAPELRADSRVNDNAFTYHNGASDVFVAVMTAPGCGAYGTGLV
    DDLNQLRGLQGVDTNSIALNRYLVNGSBSGPKYDLNNQATLNTATAPRGL
    YNDGSLTLTAYLTDKADTLARVDSVQAFQANDSEQASTLMAAGSITATRO
    VVQANRMDMLMAYGAVTVGLVTFPSMRAVLCAVLPVLTSLCEALMVALGIVK
    ATLPLVALGVIGVDYALYMSIVLQIRGASLSQAYRALFTGKVMVLTGITAI
    GVGTWIFSPKIFQADKGVLAFFVMNMVGAALILPALAVFLPHRKACTEQALPAH
    EHTRRRRSGMTEOGASFRLEEVSHGR"
    10578..11768
    /locus_tag="PP3203"
    10578..11768
    /locus_tag="PP3203"
    /note="identified by match to Pfam protein family HM
    PF04171"
    /codon_start=1
    /transl_table=11
    /product="major facilitator family transporter"
    /protein_id="AA068811.1"
    /db_xref="GI:24984835"
    /translaction="MVAERTPAGSAQAALLFGSLPVLGAVLAPVLRPMHFAEAP
    GVAVVVALTLTPALVATLALPLAGVADGRRARALLLSGLIPMLDGLV
    LTVASRAGTGLAENGIMTCCTTLMGDTFDCQRRARFALOMVTLSLAALPKVGGAL
    GESDWRTPFTLVAVGALCLPLMALMLWEPACQASBASVDSFPYMAITPLVTLV
    AGVSLPVPVQAGYVLQHLHIDAPOVGLMGANQLGLVALFRLALPGRGLLA
    MGFATIGLGGMLAASHPAVLAVINLIGVGLLPTLITLMQOVGVDFDGRATG
    GFTSAIFAGEFVSPLVLTATGAVATLPHALLVIAIGLLAPVCSLMRHRRAVTV
    AGAR"
    11926..12513
    /locus_tag="PP3204"
    11926..12513
    /locus_tag="PP3204"
    /note="identified by Glimmer2; putative"
    /codon_start=1
    /transl_table=11
    /product="hypotheetical protein"
    /protein_id="AA068812.1"
    /db_xref="GI:24984837"

Alignment Scores:
    Pred. No.:      1,71e-155      Length:      301214
    Score:          2349.00         Matches:      443
    Percent Similarity: 85.45%      Conservative: 56
    Best Local Similarity: 75.86%      Mismatches:  83
    Query Match:     76.22%         Indels:       2
    DB:              1             Gaps:         2

US-09-750-986D-30 (1-589) x AE016786 (1-301214)
QY      1 MetArgSerLeuGluAlaLeuLeuProPheProGlyVarGileLeuGluAlaArgLeuGluHis 20
Db      170844 ATGCAACCTGTGAGCACTGCGCGCTGCGCGCGCTGCTGCTGCGAGCCCTGCTGAT 170785
QY      21 TTPAlaIyethrArgProGluInthrCysValAlaAlaArgAlaAlaangIyGluTTP 40
Db      170784 TGGGCCAGGTGCGCGGACACCACTTTCATCGCGGCGAGCCGAGCAGCGGTGCTGG 170725

```

```

QY      41 ArgArgIleSerThrAlaGluMetPheHisAsnValArgAlaIleAlaInSerLeu 60
Db      170724 CGTTTCATGAGCTTACGTGAGATGCTGCGGATGTGCGACCATTCGCGGCACTTCTA 170665
QY      61 ProTyrGlyLeuSerAlaGluArgProLeuLeuIleValSerGlyAsnAerLeuGluHis 80
Db      170664 GAGTGGGCTCAGTGGCGGAGCCCGCTGCGCTGCTTCCGGAACGATGAAAC 170605
QY      81 LeuGlnLeuAlaPheGlyAlaMetTyrAlaGlyIleProTyrCysProValSerProAla 100
Db      170604 CTCCAATGCCCTCGCGCGCATGTATGCGGATATGCCATATGTCGCGGCTGCGGCG 170545
QY      101 TyrSerLeuLeuSerGlnAerLeuAlaIyLeuArgHisIleValGlyLeuLeuGlnPro 120
Db      170544 TACCGCTGTTGTCGAAAGACTTCGCCCACTTCGCCCATCTCGCGAGTGTCCACCCC 170485
QY      121 GlyLeuValPheAlaAlaAerAlaAlaProPheGlnArgAlaIleGluThrIleLeuPro 140
Db      170484 GAGTGGTCTTCTGACGAGCAGCAGCGCTTCCAGCGCGCTTCGAGCGGTGCTGAC 170425
QY      141 AspAerValProAlaIlePheThrArgGlyGluLeuAlaGlyArgArgThrValSerPhe 160
Db      170424 GATTGCGTGGCGGTGATCAAGCTGCGGCGAGTGCAGTGCAGTGCAGTGCAGTGCAGTGC 170365
QY      161 AspSerLeuLeuGlnProGlyGlyIle--GluAlaAerAlaPheAlaIleThr 179
Db      170364 GACAGCTGTTG--CAACGGATGACTTGGCGGCGCGCATTCGCGCGCAC 170308
QY      180 GlyProAerThrIleAlaIyPheLeuPheThrSerGlySerThrIyLeuProValAla 199
Db      170307 GGGCGGAGCAGCATCGCAAAATTCCTTCACTCGGCGTGCAGCAAGTGCAGCAAGCG 170248
QY      200 ValProThrThrGlnArgMetLeuCyAlaAerGlnImetLeuLeuGlnThrPhePro 219
Db      170247 GTATATCACCACCCAGGCGAGTGTGCGCAATCAGACAGATGCTTTCGAGACTTTCCG 170188
QY      220 ValPheGlyValGluProProValIleuValAerThrLeuProThrAsnIsthrPheGly 239
Db      170187 ACCTTGGCGGAGAGCGCGGCTGCTGTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTG 170128
QY      240 GlySerHisAsnIleGlyIleValIleuTyrAsnGlyGlyThrTyrTyrIleuAerAerGly 259
Db      170127 GTTAGCACAACCTCGGATCGTCTTCAACAGGGGCGAGTTTCTCACTGAGCGCGCG 170068
QY      260 LysProThrAlaGlnGlyPheAlaGluThrIleuArgAsnLeuSerGluIleSerProThr 279
Db      170067 AACCCGACCCCGAGAGCTTCCGCGAGACTTCGCAATTCGCGAGATTTCCCCACG 170008
QY      280 AlaTyrLeuThrValProLysGlyTyrGluGluValAlaGluGluAerAerSer 299
Db      170007 GCTTACCTCAGCTTACCCAGAGGCTGGAGAGACTGTCAGAGCACTGAGAGAGACCCC 169948
QY      300 ThrLeuArgGluArgPhePheAlaArgMetLysLeuPhePheAlaAlaGlyLeu 319
Db      169947 GCGCTAGCGAGGTGTTCTTTCGCCGATCAACTGTTCTTTCGCGCGCGAGGCTG 169888
QY      320 SerGlnGlyIleTyrAerArgLeuAerArgValAlaGluGlnHisCysGlyGluValIle 339
Db      169887 TCCGAAAGCCTCTCGGAGCCGCTGAGCCGCAATTCGCGAGAACTGTGCGAAGCATC 169828
QY      340 ArgMetMetAlaGlyLeuGluMetThrGluThrAlaProSerCysThrPheThrGly 359
Db      169827 CGCATAGCGCGGCTTGGCATGACAGAGCTCGCATCGTGCACCTTACACACCGGG 169768
QY      360 ProLeuSerMetAlaGlyTyrIleGlyLeuProAlaProGlyCysGluValIyLeuVal 379
Db      169767 CTTTTCGATGAGCGCGCTATGTGCGGCTGCGCGCACTGCGCAAGGAGTGTG 169708
QY      380 ProValAerGlyIyLeuGluGluArgPheHisGlyProHisValMetSerGlyTyrTTP 399
Db      169707 CCGGTGGCAGCAAGCTCGAGGGCGCTTCGTTGCGCGGATATGCGCGGCTACTGG 169648
QY      400 ArgAlaProGluGlnAsnAlaGlnAlaPheAerGluGluIyTyrTyrCysSerGlyAer 419

```

Db 169647 CGCTCCCGCAGACGCCGCGGCGCTTCAGACGAGGCGCTTACTGTCGGGCGAC 169588
Qy 420 AAla1LysLeuAlaAspProAlaAspProGlnLysGlyLeuMetPheAspGlyArgIle 439
Db 169587 GCGTTGAGCTGGCCGATGCCAGGACGCCGAGCTGGCTGATGTTGATGGCCGATC 169528
Qy 440 AlaGluAspPheLysLeuSerSerGlyValPheValSerValGlyProLeuArgThrArg 459
Db 169527 GCTGAGACTTCAAACTTTCGTCGGGGGTATTCGTAGTGTCCGGCCGCTGCCAACCCG 169468
Qy 460 AlaValLeuGluGlyGlySerTyrValLeuAspValValAlaAlaPheAspArgGlu 479
Db 169467 GCAGCTCGAGGGCTCGCCCTTACGACAGACATCGTGTCAACCGCGCGACCGGAA 169408
Qy 480 CysLeuGlyLeuLeuValPheProArgLeuLeuAspCysArgAlaLeuSerGlyLeuGly 499
Db 169407 TGCCTGGGCTGCTGTTGTTCCCGCTCTCCGAGTGTGGCGCCCTGGCCGGCTGGCA 169348
Qy 500 LysGluAlaSerAspAlaGluValLeuAlaSerGluProValArgAlaTrpPheAlaAsp 519
Db 169347 GAGAGTGCACGCGATGGCGGGGTCTGGCCACGACCGTGGCGAGTTGTTCCGTGAC 169288
Qy 520 TrpLeuLysArgLeuAsnArgGluAlaThrGlyAsnAlaSerArgIleMetTrpValGly 539
Db 169287 TGGCTGAGCGCTTGAACCGCGATGCCCAAGGACCGCCGCTATCGAATGCGTCTCG 169228
Qy 540 LeuLeuAspThrProProSerIleAspLysGlyGlyValThrAspLysSerIleAsn 559
Db 169227 CTGCTGCCGACGCGCGCTGATCGAGCGCGGTGMAATCCCGACAAAGGCTGATCAAT 169168
Qy 560 GlnArgAlaValLeuGlnTrpArgSerAlaValValAspAlaLeuTyrArgGlyGluAsp 579
Db 169167 CAGCGCGCGCTGTCAGCGCGCGCGCTCAAGTCAAGCGCGCTGTACCGTGAAGAC 169108
Qy 580 GlnSerMetLeu 583
Db 169107 CCCGACGCAATG 169096

RESULT 7
PPUS36324 4407 bp DNA linear BCT 09-JAN-2004
LOCUS Pseudomonas putida ech gene, vdh gene, fcs gene and aat gene
DEFINITION (partial).
ACCESSION AJ536324
VERSION AJ536324.1 GI:40788111
KEYWORDS aat gene; beta-ketoliase; ech gene; fcs gene; feruloyl-CoA hydratase; feruloyl-CoA synthetase; vanillin dehydrogenase; vdh gene.

SOURCE
ORGANISM Pseudomonas putida
Pseudomonas putida
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
1
REFERENCE
AUTHORS Ficca, A.G., Di Gioia, D., Barghini, P., Fava, F. and Ruzzi, M.
TITLE Identification of Pseudomonas putida strain T2 genes involved in ferulate catabolism
JOURNAL Unpublished
AUTHORS Ruzzi, M.
TITLE Direct Submission
JOURNAL Submitted (09-JAN-2003) Ruzzi M., Agrobiology and Agrochemistry, DABAC - University of Tuscia, via C. de Lellis, snc - Viterbo, 01100, ITALY

FEATURES
source
1. 4407
/organism="Pseudomonas putida"
/mol_type="genomic DNA"
/strain="T2"
/isolation_source="soil"
/db_xref="taxon:303"
/country="Italy:Latvio"
10..741

gene

CDS
/gene="ech"
10..741
/gene="ech"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="feruloyl-CoA hydratase"
/protein_id="CAD60261.1"
/db_xref="GI:40788112"
/translation="MSPTLNRMRDVLLETVEODADARVLVLNGESAMTAGMDLKEYF
REVDAPEILQRRIRRDASEMOKLIRFYSKPTIMVNGMCGGSGSPVLVADLAI
DEATRGSLSTINGITPPGNLVSAMADTVGHRSLIYITGKTFDQKAQKGLYKNSV
PLAQRLDEVLLAQDLIDKNPVLRLAKNGFRCRELTWENEDYLYALQDAEIDPE
HRAQGLKQFLDDSKIKPGLQAYKR"
782..2239
/gene="vdh"
782..2239
/gene="vdh"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="vanillin dehydrogenase"
/protein_id="CAD60262.1"
/db_xref="GI:40788113"
/translation="MDVKGALLIGKARASDGTFFERRHPVSGEVSVVAATLTAD
ADAAVEAQAAPPAAPALQPGERRARLKAADALARPDEFLAMGETAKANWGFN
VMLAANIIEAASMTQIOGSVIPSDVPSFAMALROPCEVVLGIAPMNAPIYLATRA
IAMPACGNTVVLKASEAPVHGLIGVLOAAGLDGVNVIYCNAPHPADAVVERLI
AHPAVRVNFTSGTIVGRIVGOLARHUKPVLBIIGGKAPLVLADADLEAAVADA
PBAVTRQGOICMSTERLIVDQAVADEPFAKLEKERRRLAAPDSQDAPLGLIDBGA
GRIQALVDAVAGKATLVAGSERGTSVMQPVLDHVNAGMKLYRESGFPVAVVLRG
SGDEMLALANDSEFGLSAITFSRDTQRLALAQRLSEGIHINGPTVDEAQMPPGG
VVASGYSFGSGTAIDAFQLRWVTVQHGARRHYPI"
2319..4175
/gene="fcs"
2319..4175
/gene="fcs"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="feruloyl-CoA synthetase"
/protein_id="CAD60263.1"
/db_xref="GI:40788114"
/translation="MDCELTLSVAQRAPLRARVKGIPQVQVROQGDVLYLDPEVETLDAY
PARLMERLHMAQORSQETFFVARADGQMOHITYAAMQORVRIATWILGQGLSHER
PLLILSGNDLEHVALATALLVAGIYPCVPSPAYSLSDQFKLRHIVOTLQGLVFA
DAAYRAIDAVEPESINILVGGGELAGASPTPEAVLTTVAQPCADQAFALRPDLSA
KELPTSGSKLAPKAVYTTQGMTCANQOMLQTPPEPAAVPPVINDMLPNNHTRGSHN
VGIIVLNGSTLITIDGKPTPGSPQTLNHLKDITSPAYLTVFKGMBELVKALBODRAL
CBRFQRMKLEFFPAGAGLSQALWDRIDAVAVACERIRMGAGLQMTETSPCTPTG
PLSVAGYIGLPAFGCEVKLVYVDGKLEAFRQPHVMPGYRSGSOQTEAFDEGFPYCS
GDAVNLCDPENGQGLIMFQGRYADPFKLSGVFVSGPWRTOVSSGSPVLDVVVYGP
DRECGILIFPLRPGROSLGGAQASPTKVLASPVDRCLASLATLRGATGNASV
TRPAGLAAPRPIDRGEITDKSLNQRVVLQLRHGLIDSLVAGPGLRVDAP"
4172..4407
/gene="aat"
4172..4407
/gene="aat"
/codon_start=1
/transl_table=11
/product="putative beta-ketoliase"
/protein_id="CAD60264.1"
/db_xref="GI:40788115"
/translation="MTARGQFRTPAATALLGRIPTPWVDINGAFGLVSPFDLIGIRAGR
AVLQARLADAGVDSVLGSMQADFDVILPRH"

ORIGIN
Alignment Scores:
Pred. No.: 2..56e-140 Length: 4407
Score: 2103.00 Matches: 410
Percent Similarity: 79.79% Conservative: 48
Best Local Similarity: 71.43% Mismatches: 114
Query Match: 68.23% Indels: 2

DB: 1 Gaps: 2
 US-09-750-986d-30 (1-5889) x PPU536324 (1-4407)
 QY 4 LeuGluAlaLeuLeuProPheProGlyArgIleLeuGluArgLeuGluNH1STPAlaLys 23
 2430 GTGGAAACCTGATGATGATACCGGCAACGGCTGATGAGGCGCTTGCTGATGGGCGAC 2489
 QY 24 ThrArgProGluGlnThrCysValAlaAlaArgAlaAlaAsnGlyGluTPAArgAla 43
 2490 CAGCGAGGAGGAGGAGGCTTTGTGCGCGCGCGCTGCGCGAGCGGAGAGTGGACATC 2549
 QY 44 SerTyrAlaGluMetPheHisAsnValArgAlaAlaAlaGlnSerLeuLeuProTyrGly 63
 2550 ACTTACGCGCGCTGTTGCGAGCTGTTCCAGCATGCGCACGCTGCTGCTGGCGCAAGC 2609
 QY 64 LeuSerAlaGluArgProLeuLeuIleValSerGlyAsnAspLeuGluNH1SleuGlnLeu 83
 2610 CTGAGCGCGGAAAGCGCGCTGATGATCTCTCGGCGCAACGCTTGAGCAGTGAACCTG 2669
 QY 84 AlaPheGlyAlaMetTyrAlaGlyIleProTyrCysProValSerProAlaTyrSerLeu 103
 2670 GCCACGCGCGCGCTGATGCGGATTCCTATTCGCGGATTCGCGAGGATTCACATG 2729
 QY 104 LeuSerGlnAspLeuAlaLysLeuArgHisIleValGlyLeuLeuGlnProGlyLeuVal 123
 2730 TTGTCCTCAGATTTTGGCAGAGCTGCGCATATCTCCAGACACTGCAACCGGCGCTGTC 2789
 QY 124 PheAlaAlaAspAlaAlaProPheGlnArgAlaIleGluThrIleLeuProAspVal 143
 2790 TTCGCGCGCTGATGCGCGC--TACGGAGGCGCATGATGCGGTTCGAAACCTCGATT 2846
 QY 144 ProAlaIlePheThrArgGlyGluLeuAlaGlyArgThrValSerPheAspSerLeu 163
 2847 AACCTGCTGCGCGAGCGCGAGCTGCGGCTGGCGCGAGCGCCATTCGCGCGAGT 2906
 QY 164 LeuGluGlnProGlyIleGluAlaAspAsnAlaPheAlaAlaThrGlyProAspThr 183
 2907 CTGACCAACCGTCCCAACCTGCTGCGCAGCAGCTTTTGCCTTGCGCGCGGACAGC 2966
 QY 184 IleAlaLysPheLeuPheThrSerGlySerThrLysLeuProLysAlaValProThrThr 203
 2967 CTGGCCAAAGTCTGTTACCTCGCGCTCCCAAGCTGCCCAAGGGGGTGCACACACC 3026
 QY 204 GlnArgMetLeuCysAlaAsnGlnGlnMetLeuLeuGlnThrPheProValPheGlyGlu 223
 3027 CAGGGCATGCTGTGTCCAAACGAGATGCTGTGAGAGCTTCCCGGAGTTCGCGGAA 3086
 QY 224 GluProValLeuValAspTrpLeuProTrpAsnHisThrPheGlyGlySerHisAsn 243
 3087 GTGCGCGCGGTGTGTGTCGATGCTGCGCGTGGAAACACAGTTCGCGGACGACACAC 3146
 QY 244 IleGlyIleValLeuTyrAsnGlyGlyThrTyrTyrLeuAspAspGlyLysProThrAla 263
 3147 GTCGGATGCTGTGTCAACGCGCGGAGCGCTGTACATCAACATGCGCAAGCGAGCCG 3206
 QY 264 GlnGlyPheAlaGluThrLeuArgAsnLeuSerGluIleSerProThrAlaTyrLeuThr 283
 3207 CAGGGCTTTCGCGCAACGCTGCAACCTTCAGAGCATTTCCCGCAGCGGCTGACACC 3266
 QY 284 ValProLysGlyTyrGluGluValAlaGlyAlaLeuGluArgAspSerThrLeuArgGlu 303
 3267 GTGCGCAAGGCTGGAGGAGTGTTCAGGCACTGAGAGAGAGCGCGCGCTGTGGAG 3326
 QY 304 ArgPhePheAlaArgMetLysLeuPhePheAlaAlaAlaGlyLeuSerGlnGlyLe 323
 3327 CGGTTCTTCAGGCGATGAAGCTGTTCTTTTCGCGCGTCCGGGTTGTCCAGGCGCATC 3386
 QY 324 TrpAspArgLeuAspArgValAlaGluGlnHisCysGlyGlyLysArgIleArgMetAla 343
 3387 TGGGACCGCGCTTACCGCGCTGTGGCCCACTGCGCGCAAGCATCCCGCATGATGGCG 3446
 QY 344 GlyLeuGlyMetThrGluThrAlaProSerCysThrPheThrGlyProLeuSerMet 363

DB 3447 GGGCTGGGATGACCGAGACTTGCGCGCTGTGACATTTCAACACGCGCGCTGTGCTG 3506
 QY 364 AlaGlyTyrIleGlyLeuProAlaProGlyCysGluValLysLeuValProValAspGly 383
 DB 3507 GCCGCTTACATCGCTGCGCGCGCGCGCGGTGAGAGTGAAGCTGTGTGCGGAGCGCG 3566
 QY 384 LysLeuGluGlyArgPheHisGlyProHisValMetSerGlyTyrTrpArgAlaProGlu 403
 DB 3567 AAGCTGAGGCGAGGTTTGGCGGCCCCACGTGATGCGGGGATCTGCGGCTCATGCGAG 3626
 QY 404 GlnAsnAlaGlnAlaPheAspGluGluGlyTyrTyrCysSerGlyAspAlaIleLysLeu 423
 DB 3627 CAGACCCAGAGGCGTTCCAGCAGAGAGGGGTTCTACTGACGCGGTATGCGGCGCTG 3686
 QY 424 AlaAspProAlaAspProGlnLysGlyLeuMetPheAspGlyArgIleAlaGluAspPhe 443
 DB 3687 TGGACCCCGCGCAACCGGAGCTGCGTTGATGATTTGACGCGGCGCATTCGCGAGACTTC 3746
 QY 444 LysLeuSerSerGlyValPheValSerValGlyProLeuArgThrArgAlaValLeuGlu 463
 DB 3747 AAGTTTCTTCGCGGAGTTTGTGACGCTTGCGCGCATGCGCACCCAG--GTTCTTCA 3803
 QY 464 GlyGlySerTyrValLeuAspValValAlaAlaProAspArgLysLeuGlyLeu 483
 DB 3804 GGGTCCGCTATGCTGATGATGTGTGTGGGCGCGGACCGGAAAGCTGGGTTTG 3863
 QY 484 LeuValPheProArgLeuLeuAspCysArgAlaLeuSerGlyLeuGlyLysGluAlaSer 503
 DB 3864 CTATATCTTCGCGGCTTGGCGGAAATCCGCGAGTTTCTCGGCGCTTGGCGCGCGCACT 3923
 QY 504 AspAlaGluValLeuAlaSerGluProValArgAlaAlaTrpPheAlaAspTrpLeuLys 523
 DB 3924 ACCGCCAAGGTGTGCAATGCGCGCGCTGCGAGTCTGCTGCGCGCTGCGCGCAGC 3983
 QY 524 LeuAsnArgLysAlaThrGlyAsnAlaSerArgIleMetTyrValGlyLeuLeuAspThr 543
 DB 3984 CTCAACCGGCGCGCGAGGCAAGCGCACGCTACATCGCTGCGCGCGCGCGCGGAA 4043
 QY 544 ProProSerIleAspLysGlyGluValThrAspLysGlySerIleAsnGlnAlaVal 563
 DB 4044 GCACCGCGCATCAGCGCTGTGTGATGATCAGCAGCGGCTGCTCAACGCGCGTGTGTC 4103
 QY 564 LeuGlnTrpArgSerAlaLysValAspAlaLeuTyrArgGly 577
 DB 4104 CTGCAATTCGGGACGCGCTGATGACAGCTGTATGCAAGG 4145
 RESULT 8
 AL646077 189050 bp DNA linear BCT 11-Apr-2003
 LOCUS
 DEFINITION
 Ralstonia solanacearum GMI1000 megaplasmid, complete sequence;
 segment 2/11.
 ACCESSION
 AL646077 AL646053
 VERSION
 AL646077.1 GI:17430642
 KEYWORDS
 SOURCE
 ORGANISM
 Ralstonia solanacearum
 Ralstonia solanacearum
 Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 Burkholderiaceae; Ralstonia.
 REFERENCE
 1
 AUTHOR
 Salanoubat, M., Genin, S., Artiguenave, F., Gouzy, J., Mangenot, S.,
 Arlat, M., Billault, A., Brotier, P., Camus, J. C., Cartolico, L.,
 Chandler, M., Choisme, N., Claudei-Renard, C., Cumac, S., Demange, N.,
 Gaspin, C., Lave, M., Moisan, A., Robert, C., Saurin, W., Schlex, T.,
 Siguer, P., Thebaud, P., Whalen, M., Wincker, P., Levy, M.,
 Weisenbach, J. and Boucher, C. A.
 TITLE
 Genome sequence of the plant pathogen Ralstonia solanacearum
 Nature 415 (6871), 497-502 (2002)
 JOURNAL
 MEDLINE
 PUBMED
 21681879
 11823852
 2 (bases 1 to 189050)
 REFERENCE
 Boucher, C. A.
 TITLE
 Direct Submission

JOURNAL Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie
Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,
BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
Dausset-CEPH, 27 rue Juliette Bodu, 75010 Paris, France, LMGM CNRS
118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA
URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,
Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,
F31326 Castanet-Tolosan Cedex, Laboratoire de Genetique Cellulaire
INRA, BP27, F31326 Castanet-Tolosan Cedex
COMMENT
http://sequence.toulouse.inra.fr/R.solanacearum.html.
location/Qualifiers
1. 189050
/organism="Ralstonia solanacearum"
/mol_type="genomic DNA"
/strain="GMI1000"
/db_xref="taxon:305"
/plasmid="megaplasmid"
79. 1500
/note="R49+ or RS06146
LONG DIRECT REPEAT 49 +
predicted by Homology"
/evidence=not_experimental
83. 2971
/gene="RSp0175"
/note="synonym: RS04687"
83. 2971
/gene="RSp0175"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PUTATIVE VGR-RELATED PROTEIN"
/protein_id="CAD17326.1"
/db_xref="GI:17430643"
/db_xref="GOA:Q8XTD8"
/db_xref="TrEMBL:Q8XTD8"
/translation="MDFKTLQHLFAPARHLYALBEGCPILAEIYAVEMILGREALSELF
EMRVASANARFALSPFGORVTLVTLADGQARGTGLIROEQLGADGSLARYL
TVVPMWLTQOHSQVFNORPLADILEQVLSFARYAAMRPAAGADEMAARGTSH
IAOFRTDYRFRTRILAEAGCTGTTEDEQAPSGHLLIFADSPRLAEDTASAEGL
RTHRASQSEBDALIQALICHSRTAVGVVAAMDQAQKAFRAHVSPRCGLASGSDA
YLSISPLAPDVTNAQRILAEQVESVEARLPLFIRGSVTLASGTRLRIVDCPLPK
DVDPYPLLDIVEHGGINLLTADQATLARKLGIDALITPDTPSPSPSGGPFPG
MAPHEVIERFTPADLLTARAHGAVGQFRAGDARPMRPVSSHPTGRLVAEPTAG
VGSATVVGNGETFEASGDAEHHTTPRGQVVRPMQGRPDPRSTGMILVQROGA
GNGKWLPRIGQEVLYRPOEDDIDQPVVIGALYNGGEGAIPTPGQVYAKHROGAD
PELTVLGSDSASAOGNLAHSPAMHGIGTEBGRNNAALISFKSAHGGKGNL
LVFDDSDQLRQLATTOAYSQLNLGHLHQDNRRGSPRGQFELRTDGYGVRQSA
GVLLTYRDATSGKAVPTGDNAGAILIRQAKQLTRLSQGATHTQALSTAKDQA
PLAKOKAASGMWDKALDAODDAAGNVTATQGVPHOSEATVVLQAGAGVAILQGO
DLOFANGESVILASGODTVNAVQKARHAGGIGVAGLSTAGDNDNLGLQITAGQDN
LDYQAOHDALGLSKDDKLVSKNLHVDPAAARIRLATAGASTLBEGNITVETPG
RITYKAQRSFAPVTSYPLVPFGQVCEVLRAMGKOHAFANNG"
1501. 2865
/note="R450+ or RS06148
PROBABLE LONG DIRECT REPEAT 50 +
predicted by Homology"
/evidence=not_experimental
2999. 3835
/gene="RSp0176"
/note="synonym: RS04688"
2999. 3835
/gene="RSp0176"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical

predicted by Codon usage
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAD17327.1"
/db_xref="GI:17430644"
/db_xref="GOA:Q8XTD7"
/db_xref="TrEMBL:Q8XTD7"
/translation="MATRFQIDDPHRRPRAWVDYPADAVEQLLSALPRLDYALVDN
AYDPARHTRRRPDPMPRLXARGVSGDELATPRLMDPAABERRTLDLFLQ
ETSGKPMISFTLHANAADPRLTHLCKQHEAVDHEGTAFILRLADTALDAIVQVND
QORFLSDTQWCFRRDSLRHSAHAGGSGAESTPYKCTAQVQLEALRPGAML
NLVQNSPHIFGDLTGLPSQAYECIRMALQALRPGNPTDAVALRLIASALNAGLLQ
AA"
4412. 6634
/gene="RSp0177"
/note="synonym: RS04689"
4412. 6634
/gene="RSp0177"
/function="miscellaneous; hypothetical/partial homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAD17328.1"
/db_xref="GI:17430645"
/db_xref="GOA:Q8XTD6"
/db_xref="TrEMBL:Q8XTD6"
/translation="NSVLEPQNPALTYVDQARFLACRMQEFERHCSTERHLSFF
DSTNNRRYKDTPRQASHSVARLPDIPEDEHGRIVPEIGPFEKEDTGRGDAR
AGLGGWGGEARINNALQVTLALAYYTOGLSMAGVDELSLVHQTSNLMGLGA
VTTAGRDEVEOLANAKNSEMYTAAETVTLPRHBERVALLQREBEYSEKILKALIG
RKRMILCTRLSVGFSGRGAELARVANNMKDLDDMTLAGVAFQDFIGIPDYVAS
GVANSTKYATGHSFGGEAEFLRIPGVVRYTLVLSAHEHRSSEPLDKVANDCLEY
PEVHTDVGAIVPGQGRCCGTDGKRPDSNKLSQTLAMRYEAAAAGPLPMMARNL
TEVFRALKISPDLLQAYNDVDVAVNPLIRKGGTTCGARVQYGLYLKMRWRLLGG
AQAENQLPFGQAQVGAQCANDLVANALLKEAKDLAARENDPAVYDGMWRQVR
LEVTOGVAVNMRQWGDKYREVRVAVYANDPDLPRIVRLICDDYIHDSRAMFK
PGAPSDSVNMRQOARLEOLKOQAAWQVADISRLIGTLKYEYSTANGDGSAA
EMPVVIGORRDLQOYLKXGQVPMETKGRSPSMGVIARWTHFAPLPTLIGERAQAMD
EVASVPGYAKKAQZAAHDADEARLANTARKLLEAGQESLERRATTAQLRFLGGYPRP
"
6745. 7527
/gene="RSp0178"
/note="synonym: RS04690"
6745. 7527
/gene="RSp0178"
/function="miscellaneous; unknown"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAD17329.1"
/db_xref="GI:17430646"
/db_xref="GOA:Q8XTD5"
/db_xref="TrEMBL:Q8XTD5"
/translation="MTQYLYHTTTAVARILRTKGLTLAAHPEALGPVARRHGAFEV
NRAAQEPGQVNRLLRAYLKKGLGAGYSIDQIRTGORPFTPIPVVPAGRNDDQVEITR
VBEAEVKAFLAIGTPANKDGRLTWDLTTEGHADMDLRTRRANALGRAYHVSLEY
AIEEGMTSHVTFSPERASDCYSYTRQHGGAACSVLRVSRMAAALPLDDPSDFRA
VWTQRRILLPQQLIWRAPSDVLTINADPRAAGNMPLTQWS"
7518. 8021

/gene="RSP0179"
/note="synonym: RS04691"
7518. 8021
/gene="RSP0179"
/function="miscellaneous; hypothetical/partial homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon usage
predicted by FrameD
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAD17330.1"
/db_xref="GI:17430647"
/db_xref="GDA:Q8XTD4"
/db_xref="TrEMBL:Q8XTD4"
/translation="MELNADRMGQPTMRAVGRGAALIALSALALAPALADAPKPMQ
DEPARAVSEAOPEVVSSTHPLIVLIHADPAMVAQSTVDVDRHNVALPKGSRLI
GKEIRVNAAREIIVWIGLOIPASGTRLRDPLQATMRDSAGVADLSFGALLGAMTS
EPRTVPH"

gene
complement (8066. 8488)
/gene="RSP0180"
/note="synonym: RS04692"
complement (8066. 8488)
/gene="RSP0180"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon_usage

Alignment Scores:
Pred. No.: 1.39e-126 Length: 189050
Score: 1940.50 Matches: 373
Best Similarity: 75.43% Conservative: 69
Best Local Similarity: 63.65% Mismatches: 143
Query Match: 62.96% Indels: 1
Gaps: 1

US-09-750-986d-30 (1-589) x AL646077 (1-189050)

Qy 1 MetArgSerLeuGluAlaLeuLeuProPheProGluYarGileLeuGluArgLeuGluHis 20
Db 66491 CTGGCTCGGCTGAGCCGATCGTGTCTATCCGAGCGATGACGATGCGTGTGCGC 66550
Qy 21 TrpAlaYstHrArgProGluGlnThrCysValAlaAlaArgAlaAlaAsnGlyGluTrp 40
Db 66551 GGGGCGGACAGATCCGATCCGCTCGCGCGCGAGCGCGCGCGATGCGGCACTGG 66610
Qy 41 ArgArgIleSerYrAlaGluMetPheHisAsnValArgAlaIleAlaGlnSerLeuLeu 60
Db 66611 GAGGCGATCACTATGCGAGATCTGTACCGGCGCGCGCGCGCGCGCGCGCTCTG 66670
Qy 61 ProTyrGlyLeuSerAlaGluArgProLeuLeuIleValSerGlyAsnAspLeuGluHis 80
Db 66671 GCGCGGGGGGCTGTGCGCGAGCGCGCTGTCTATCTCTGCGGCAACGATCTCAAGAT 66730
Qy 81 LeuGlnLeuAlaPheGlyAlaMetYrAlaGlyIleProTyrCysProValSerProAla 100
Db 66731 CTGAGCTGGGCGTGGGCGCATATGCGGCACTTCCCTATGCCCCGCTGCGCGGCC 66790
Qy 101 TyrSerLeuLeuSerGlnAspLeuAlaValLeuArgHisIleValGlyLeuLeuGlnPro 120
Db 66791 TACCGCTGTGAGCGAGCACTACAGCGCGCTCTCCCTACGTCGTGCGGCACTTACCGCG 66850
Qy 121 GlyLeuValPheAlaAlaAspAlaAlaProPheGlnArgAlaIleGlnThrIleLeuPro 140
Db 66851 GGCGTGGCTATGCAACGAGCGTGGCTGTTCGCGCGCGCGCATCCAGCGCGTGTGCGG 66910
Qy 141 AspAspValProAlaIlePheThrArgGlyGluLeuAlaGlyArgArgThrValSerPhe 160
Db 66911 CCGGAGACCGAGCTCGGATGACCAACGCGAGGTCGGGCGCGCGCGTGAACCGGCTG 66970

Qy 161 AspSerLeuLeuGluGlnProGlyGlyIleGluAlaAspAsnAlaPheAlaIleThrGly 180
Db 66971 GCCTCGCTGCTC---GGAACCGGGCGGCGCGATCGATCGCCCAACCGCGCGTGGCG 67027
Qy 181 ProAspThrIleAlaYstPheLeuPheThrSerGlySerThrIstLeuProValAlaVal 200
Db 67028 CCGGACAGATCCCGCAAGTTCCTGACCTCGGGCTCGACCGCGGAGCCCAAGGGCGTCC 67087
Qy 201 ProThrThrGlnArgMetLeuCysAlaAsnGlnGluMetLeuLeuGlnThrPheProVal 220
Db 67088 ACCACACGATGATGATGCTGTGCGACGACACACAGATGCTGTCGACACTTTCCTGCG 67147
Qy 221 PheGlyGluGluProProValLeuValAspTyrLeuProTyrAsnHisThrPheGlyGly 240
Db 67148 TTGCGGAGGCGCGCGCGCTGCTGCTGATGCTGCTGCGCGGACCAACACTTCGCGCGC 67207
Qy 241 SerHisAsnIleGlyIleValLeuTyrAsnGlyGlyThrTyrTyrLeuAspAspGlyLys 260
Db 67208 AGCCACAACTGGGACCTCGCGCTGTCAACGCGGACGACTACTACATCGACAGCGCAG 67267
Qy 261 ProThrAlaGlnGlyPheAlaGluThrLeuArgAsnLeuSerGluIleSerProThrAla 280
Db 67268 CCGACACCGGAGCGCTTCGACGACGCTGCGAACCCTGCGCAACGCGACCGACGCTG 67327
Qy 281 TyrLeuThrValProIstGlyTyrGluGluGluValGlyAlaLeuGluArgAspSerThr 300
Db 67328 TACTTCAACTGCTCCCAAGGCTGGAGATCTACACGACCGCTCGACGCGGACCCGCC 67387
Qy 301 LeuArgGluArgPhePheAlaArgMetLysLeuPhePheAlaAlaAlaGlyLeuSer 320
Db 67388 CTGCGGACGACTTATGCGCGCGCTGCTGCTCTTCTTCTGCGGCGCGGCGCTGCTCG 67447
Qy 321 GlnGlyIleThrAspArgLeuAspArgValAlaGlnGlnHisIstCysGlyGluArgIleArg 340
Db 67448 CAGCGCGCGGAGCGGCTCTGAGGCGGTGACCGAGCCCACTGCGCGGCGCGATTCGC 67507
Qy 341 MetMetAlaGlyLeuGlyMetThrGluThrAlaProSerCysThrPheThrThrGlyPro 360
Db 67508 ATCATGCGCGGCTGGGAGTACCGAGACCGCGCGCTGTCGATGATCAACACCGGCGC 67567
Qy 361 LeuSerMetAlaGlyTyrIleGlyLeuProAlaProGlyCysGluValLysLeuValPro 380
Db 67568 GTCATATGCGCGGCTCATCGGCTGCGCGCGCGCGCGCGCGCGCAAGCTCGTGGCG 67627
Qy 381 ValAspGlyLysLeuGluGlyArgPheHisGlyProHisIstAlaMetSerGlyTyrTrpArg 400
Db 67628 GTGACCGGCAAGCTCGAAGCCGCTTCMAAGGCGCGCGCATGTCATGCGGCTACTGGCGC 67687
Qy 401 AlaProGluGlnAsnAlaGlnAlaPheAspGluGluGlyTyrTyrCysSerGlyAspAla 420
Db 67688 GCGCGGAGCTGGCGCTGAAAGCTTGACGAAAGAGCGCTACACCGCTCGGCGGAGCGC 67747
Qy 421 IleYstLeuAlaAspProAlaAspProGlnLysGlyLeuMetPheAspGlyArgIleAla 440
Db 67748 GTGGCTTGTGACGAGCGGCGCGCGCGAGATCGGCTGATGTTCAGCGCGCGCATGCC 67807
Qy 441 GluAspPheLysLeuSerSerGlyValPheValSerValGlyProLeuArgThrArgAla 460
Db 67808 GAGGACTTCAAGCTCAAGCTCGGACCTTGTGAGCTGTGGCGCGCTGCGCGCGCGGATC 67867
Qy 461 ValLeuGluGlyGlySerTyrValIleAspValValAlaAlaProAspArgLysCys 480
Db 67868 ATCAGCGAGGCGCGCGCTTACGCTGCTGACGCGGTATCCCGGAGATGAACCGGAGAC 67927
Qy 481 LeuGlyLeuLeuValPheProArgLeuLeuAspCysArgAlaLeuSerGlyLeuGlyLys 500
Db 67928 CTCAGCGTGTGGTGTTCGCGCGCTGAGACATGCGCGCGCTGCGGCGCTGCGGCGCC 67987
Qy 501 GluAlaSerAspAlaGluValIleAlaSerGluProValAlaArgAlaThrPheAlaAspTyr 520
Db 67988 GCGCGCGCGCGCGCGAGTCCGACGACCGCTGCGGTGGCGGCGCTTCTGCGGCTGCTG 68047
Qy 521 LeuLysArgLeuAsnArgGluAlaThrGlyAsnAlaSerArgIleMetTyrValGlyLeu 540

OY		503	Serznpaiagluvalleualsersclunpro[al]arglaatpphealaastrpleunls	522
Db		1630	TCTGAGCAGAGATGTGCACCAACCCTTAAAGTCGAATGGTTCCGCCAATTTTTAACG	1689
OY		523	Argleubena[vgglua]lathrglyabna[asera]gilemetrpaiglyeuleunsp	542
Db		1690	ACCTTTAATAAGATGCACCTCGCAGCTCCAATACAGTCTCAAGCTTATTATTATGACC	1749
OY		543	ThrProPserSrlieaspLygIygluValThirapLygsrlieasnglnarGla	562
Db		1750	GAGCACCTCAGTTAGATGATCGCGCAATGACCGATTAAGCAACTCAATCAAGAAGCT	1809
OY		563	Valleugnt[rpa]argserAlalyValAspalaleutrya[rgly]giuasglnsermet	582
Db		1810	ATTACCAAGCCGTGGCGCTTTAATTGATGAGCTTTATCAAAAAGCAGACTGATTAACCG	1869
OY		583	Leu 583	
Db		1870	CTG 1872	
RESULT 10				
LOCUS	ACCPCAOP	45324 bp	DNA	linear BCT 27-JUN-2003
DEFINITION	Acinetobacter sp. ADPI dea-pca-gui-pob-hca supraoperonic gene cluster, complete sequence.			
ACCESSION	L05770 AY056039 L03407 J13114 M33798 U04359 U1554 U20284			
VERSION	L05770.5 GI:32306886			
KEYWORDS	.			
SOURCE	Acinetobacter sp. ADPI			
ORGANISM	Acinetobacter sp. ADPI Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Moraxellaceae; Acinetobacter.			
REFERENCE	1 (bases 1 to 45324) Hartnett,C., Neidle,E.L., Ngai,K.L. and Ornstom,L.N. DNA sequences of genes encoding Acinetobacter calcoacetius protocatechuate 3,4-dioxygenase: evidence indicating shuffling of genes and of DNA sequences within genes during their evolutionary divergence			
AUTHORS	J. Bacteriol. 172 (2), 956-966 (1990)			
TITLE	90130333 2298704			
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE	2 (bases 1 to 45324) DiMarco,A.A., Averhoff,B.A., Kim,E.E. and Ornstom,L.N. Evolutionary divergence of pobA, the structural gene encoding p-hydroxybenzoate hydroxylase in an Acinetobacter calcoacetius strain well-suited for genetic analysis			
AUTHORS	Gene 125 (1), 25-33 (1993)			
TITLE	93194074 8449410			
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE	3 (bases 1 to 45324) DiMarco,A.A., Averhoff,B. and Ornstom,L.N. Identification of the transcriptional activator pobR and characterization of its role in the expression of pobA, the structural gene for p-hydroxybenzoate hydroxylase in Acinetobacter calcoacetius			
AUTHORS	J. Bacteriol. 175 (14), 4499-4506 (1993)			
TITLE	93322329 8331077			
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE	4 (bases 1 to 45324) Hartnett,G.B. and Ornstom,L.N. Acquisition of apparent DNA slippage structures during extensive evolutionary divergence of pcbd and catc genes encoding identical catalytic activities in Acinetobacter calcoacetius			
AUTHORS	Gene 142 (1), 23-29 (1994)			
TITLE	94237485 8181753			
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE	5 (bases 1 to 45324) Kowalchuk,G.A., Hartnett,G.B., Benson,A., Houghton,J.E., Ngai,K.L. and Ornstom,L.N. Contrasting patterns of evolutionary divergence within the Acinetobacter calcoacetius pca operon			
AUTHORS	Gene 146 (1), 23-30 (1994)			
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE	94341565			

PUBMED REFERENCE AUTHORS TITLE	8063101 6 (bases 1 to 45324) Elsemore,D.A. and Ornstom,L.N. The pca-pob supraoperonic cluster of Acinetobacter calcoaceticus contains gula, the structural gene for guinate-shikimate dehydrogenase
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	J. Bacteriol. 176 (24), 7659-7666 (1994) 95095936 8002591 7 (bases 23320 to 27131) Elsemore,D.A. and Ornstom,L.N. Unusual ancestry of dehydrogenases associated with guinate catabolism in Acinetobacter calcoaceticus
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	J. Bacteriol. 177 (20), 5971-5978 (1995) 96011389 7592351 8 (bases 1 to 45324) Gerischer,U., Segura,A. and Ornstom,L.N. PCU, a transcriptional activator of genes for protocatechnate utilization in Acinetobacter
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	J. Bacteriol. 180 (6), 1512-1524 (1998) 98175676 9515921 9 (bases 1 to 9706) Parke,D., Garcia,M.A. and Ornstom,L.N. Cloning and genetic characterization of dca genes required for beta-oxidation of straight-chain dicarboxylic acids in Acinetobacter sp. strain ADP1
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	Appl. Environ. Microbiol. 67 (10), 4817-4827 (2001) 21455042 11571189 10 (bases 1 to 45324) Parke,D. and Ornstom,L.N. Hydroxycinnamate (hca) catabolic genes from Acinetobacter sp. strain ADP1 are repressed by HcaR and induced by hydroxycinnamoyl-CoA thioesters
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	Appl. Environ. Microbiol. (2003) In press 11 (bases 1 to 45324) Smith,M.A., Weaver,V.B., Young,D.M. and Ornstom,L.N. Genes for chlorogenate and hydroxycinnamate catabolism (hca) are linked to functionally related genes in the dca-pca-gui-pob-hca chromosomal cluster of Acinetobacter sp. strain ADP1
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	Appl. Environ. Microbiol. 69 (1), 524-532 (2003) 22401456 12514037 12 (bases 12470 to 15266) Gerischer,U.C. Direct Submission
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	Submitted (15-DEC-1993) Biology, Yale University, 266 Whitney Avenue, New Haven, CT 06520-8103, USA 13 (bases 27214 to 29733) Elsemore,D.A. Direct Submission
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	Submitted (28-JUN-1994) Biology, Yale University, 266 Whitney Avenue, New Haven, CT 06520-8103, USA 14 (bases 23320 to 27131) Elsemore,D.A. Direct Submission
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	Submitted (25-JUN-1995) Biology, Yale University, 266 Whitney Avenue, New Haven, CT 06520-8103, USA 15 (bases 1 to 45324) Ornstom,L.N. Direct Submission
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	Submitted (02-JUN-1998) Biology, Yale University, 266 Whitney Avenue, New Haven, CT 06520-8103, USA 16 (bases 1 to 45324) Ornstom,L.N. Direct Submission
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	Submitted (20-FEB-2001) Biology, Yale University, 266 Whitney Avenue, New Haven, CT 06520-8103, USA 17 (bases 1 to 45324) Sequence update by submitter

AUTHORS Smith,M.A., Young,D.M. and Ornston,L.N.
TITLE Direct Submission
JOURNAL Submitted (12-SEP-2001) MCDB, Yale University, P.O. Box 208103, New Haven, CT 06520-8103, USA
REFERENCE 18 (bases 1 to 45324)
AUTHORS Smith,M.A., Young,D.M. and Ornston,L.N.
TITLE Direct Submission
JOURNAL Submitted (07-MAY-2002) MCDB, Yale University, P.O. Box 208103, New Haven, CT 06520-8103, USA
REMARK Sequence update by submitter
AUTHORS 19 (bases 1 to 45324)
TITLE Direct Submission
JOURNAL Submitted (27-JUN-2003) MCDB, Yale University, P.O. Box 208103, New Haven, CT 06520-8103, USA
REMARK Sequence update by submitter
COMMENT On Jun 27, 2003 this sequence version replaced gi:20492839.
FEATURES
source
1. .45324
/organism="Acinetobacter sp. ADPI"
/mol_type="genomic DNA"
/strain="ADPI"
/db_xref="taxon:62977"
/note="formerly Acinetobacter calcoaceticus"
/complement (<1. .689)
/gene="dcaF"
/complement (<1. .689)
/gene="dcaF"
/function="putative thiolase in catabolism of dicarboxylic acids"
/codon_start=1
/transl_table=11
/product="DcaF"
/protein_id="FAL09090.1"
/db_xref="GI:15812038"
/translation="MNAHYIDGLRTPFGHAGELASIRPDDAGIVYIQRILKRTVA
GAIEDVIFEDTIOAGEDSRNIAHALLAGLVTVPGQIVNRLCASGLAIIIDSARA
ITCGEDLIYAGVESMSRAPFVWGAAESYSDAKIYDTTIGTRPNKKIYAQYGGH
SMETGDNVAVEGYISREQADLFPAOSQAKYQALREGFAGEITFAVEVSQKKLPK
QVTEDEHPRPS"
/complement (710. .2227)
/gene="dcaH"
/complement (710. .2227)
/gene="dcaH"
/function="hydroxacyl-CoA dehydrogenase in catabolism of dicarboxylic acids"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="DcaH"
/protein_id="FAL09091.1"
/db_xref="GI:15812039"
/translation="MTHPIKTAIIIGVWGSGIAQIAOSGHITLYLDAKAGAAOA
KOGLATTPQKLDKPKITTEYADPAANALLIANELDKQCDLIVAIYERLDIKSL
MSGLEIVPEPTTILASNTSLITALASNCKPERVAGHFENVPDMKVEYDLSLK
TDPKHEITLNQSRVLAHSPVAKDTPGFIINAGRAYGEALKIINENYTDISEDR
IILDVGFRMGPELMDLIGLDVSHPMESIYHYYEBAARYNSLTOKMLEAKQGR
KVGQGFYRTGKTGETSAKVAERLLTVPKWIAADFEDDKOLLINYLTHNIQD
GAKPOADICILACGEDETHALRLINVAHSAIIDLVIKIEKHTTMSITTEY
SHAHSIFNIDGAMVSTIGESIGFVAQRIIAWYINIGCDIAQQAISVDDINAAVGLG
LGYFPIEWGDISNKTLLINRIITALTSDFRPSFWLQRRVALNLPVITTT"
/complement (2276. .3037)
/gene="dcaC"
/complement (2276. .3037)
/gene="dcaC"
/note="putative ketoacyl-CoA reductase"
/codon_start=1
/transl_table=11
/product="DcaC"
/protein_id="FAL09092.1"
/db_xref="GI:15812040"
/translation="MARRKYLITAGSGIGRCIAEALIKQCDQVFCVDINDSHIQORR
DYPQLHTFCDDIGQPKRYVQAMFDYAVQQLGGDILLINNTGVSGETVAABEELSFEEWTQ

VINLINSTFVTKLAIPYLKOSDAGVILNLSIAGRMGYPRRLAYSTKMGIVGFK
TLMSEIGADIRVNAILPGAVDERFORLQARSMNISIDVTENALANOSIKHVF
NPKHIALDLCIFIASDSGRISGQIIPIDDKORLT"
Alignment Scores:
Pred. No.: 1,31e-104 Length: 45324
Score: 1621.50 Matches: 308
Percent Similarity: 69.54% Conservative: 96
Best Local Similarity: 53.01% Mismatches: 174
Query Match: 52.61% Indels: 3
DB: 1 Gaps: 3
US-09-750-986D-30 (1-589) x ACCPCAP (1-45324)
Qy 5 GUAAlaLeuLeuProPheProGlyArgIleLeuGluArgLeuGluIuhIstrPaIatyrh 24
Db 39875 GAACAACTCAAGCCTTATCCGAAACACGACCGATCGTTTGAATTGCTAAAGTT 39816
Qy 25 ArgProGluGlnThrCysValAlaAlaArgAlaAlaAsnGlyGluTrpArgArgIleSer 44
Db 39815 AAGCCCTACCAATACCTTTCGCCCTMAACCGCAATACACAAGACGAATGATCAAGCTCAGT 39756
Qy 45 TyrAlaGluMetPheHisAsnValArgAlaIleAlaGlnSerIleuLeuProTyrGlyLeu 64
Db 39755 TATGACGAAACCTTGCAACGCTGCTGGCATATTGACAAAGCTTGATACGCCAACTTA 39696
Qy 65 SerAlaGluArgProLeuLeuIleValSerGlyAsnAspLeuGluIuhIseuGlnLeuAla 84
Db 39695 AGCCAGAAACGACCTTGGTAATGTGATCGGAAAGATCTTGACACTTAACGCTTGG 39636
Qy 85 PheGlyAlaMetCysTrpAlaGlyIleProTyrCysProValSerProAlaTyrSerLeuLeu 104
Db 39635 ATGGGCGCATGTGTGCAGGTGATGACATTTTGTCTATTTCGACAGCTGATCACTCAGTA 39576
Qy 105 SerGlnAspLeuAlaIysLeuArgHisIleValIleGlyLeuGlnInProGlyLeuValPhe 124
Db 39575 TCTCAAGATTTTGGCAATTAATAACATGTGTTGATGTCTCACCCAGCGATGCTGAT 39516
Qy 125 AlaAlaAspAlaAlaProPheGlnArgAlaIleGluThrIleLeuProAspAspValPro 144
Db 39515 GCAAATGATGACACACTTTTGGCTAAAGCATTCMAAGGTGACGCAATCATATTGAA 39456
Qy 145 AlaIlePheThrArgGlyGluLeuAlaGlyArgArgThValSerPheAspSerLeuLeu 164
Db 39455 ATGTGACCAATTAAGCATTTATAGGTGATCAAGCTGTACATCGTTTCACTTATTATTA 39396
Qy 165 GluGlnProGlyGlyIleGluIaAspAsnAlaPheAlaIleThrGlyProAspThrIle 184
Db 39395 GATACCCCT--GTGACCGATGTTCAAGAACATTATGAAACGCTAGATGACATCGATT 39339
Qy 185 AlaIysPheLeuPheThrSerGlySerThrIlySLeuProIysAlaValProThrThrgln 204
Db 39338 GCCAATTTTATTATTCATCCGGTTGACCAAGCTTCCAAAGCGCTTCCGACACACAT 39279
Qy 205 ArgMetLeuCysAlaAsnGlnInMetLeuLeuGlnThrPheProValPheGlyGluGlu 224
Db 39278 TTGATCTCTGTATCATATCAACAACAACTTCAACAACTTCCAAATTTGAAGCACCA 39219
Qy 225 ProProValLeuValAspTrpLeuProTyrPasnHisThrPheGlyGlySerHisAsnIle 244
Db 39218 CGCGCTGTCTTACGACGTGGTGTCTTGGCATCATACATTCCGTGGAGCCCAATGTT 39159
Qy 245 GlyIleValLeuTyrAsnGlyGlyThrTyrTyrLeuAspAspGlyIysProThrAlaGln 264
Db 39158 GGGATTGGCGCTTATTAACGGCGGAACATTTCATCATGATGACGGTAAAGCCCGTCCGGGA 39099
Qy 265 GlyPheAlaGluThrLeuArgAsnLeuSerGluIleSerProThrAlaTyrLeuThrVal 284
Db 39098 AATTTGATGAACACGATTCGAATTTAAAGAAATTTCCACCGCGGCTTATCATTAATGTA 39039
Qy 285 ProIysGlyTTrpGluGluLeuValGlyAlaLeuGluArgAspSerThrLeuArgGluArg 304
Db 39038 CCCAAAGGCTGGGAACACTCACCGATGCACTTGAAAGATGCGCAACTCAAGATCGC 38979


```

QY 305 PhePheAlaArgMetLysLeuPhePheAlaAlaIleuSerGlnGlyIleTyr 324
Db 38978 TTCTTTCCCAAGCTTAATTTTGTTCCTGCTGCTGATATCTGAAGCAGCTGG 38919
QY 325 AspArgLeuAspArgValAlaGlnGlnHisCysGlyGluArgIleArgMetMetAlaGly 344
Db 38918 AACGCACTCGATTAATTTGACAGGACATTGGCGGTGAAAAATCCGATCATGAGTGGC 38859
QY 345 LeuGlyMetThrGluThrAlaProSerCysThrPheThrGlyProLeuSerMetAla 364
Db 38858 TTGGGCGATGACCGAAGCGGACCATCGTCCCTTTCACAGGCGCCACGCGGTATGGCT 38799
QY 365 GlyTyrIleGlyLeuProAlaProGlyCysGluValLysLeuValProValAspGlyLys 384
Db 38798 GGTATTATTTGGCTACCTGCTCCGGGCTGTGAATTCAAATTAATGTTCCCTATGTGACAAA 38739
QY 385 LeuGlnGlyArgPheHisGlyProHisValMetSerGlyTyrTyrArg---AlaProGlu 403
Db 38738 CTGGAATTTTGTGTGTCGTGGACAGCATGTAAGATTACTGGCCGCTTAAGCCGAC 38679
QY 404 GlnAsnAlaGlnAlaPheAspGluGlnGlyTyrTyrCysSerGlyAspAlaIleLysLeu 423
Db 38678 CACCAAAAGCAATTTTGTGATGAAGACGCGCTTTTCCATCTGCGCATGCGGTACGTCTG 38619
QY 424 AlaAspProAlaAspProGlnLysGlyLeuMetPheAspGlyArgIleAlaAspPhe 443
Db 38618 GTGGATCAACAGCCCTGACAGGCGCTGATGACAGCGCCGTATTTGCAAGAAATTTC 38559
QY 444 LysLeuSerSerGlyValPheValSerValGlyProLeuArgThrArgAlaValLeuGlu 463
Db 38558 AAATCAACAGCGGTACTTTGTGAATGTGGGCACTTACCAACAAAGCATCTATCCA 38499
QY 464 GlyGlySerTyrValLeuAspValValValAlaAlaProAspArgGluCysLeuGlyLeu 483
Db 38498 GGTAAATTTGTGATTCAAAGATGTTTGTATTAACCGGTTCAAATCAATGCGATGGCTTC 38439
QY 484 LeuValPheProArgLeuLeuAspCysArgAlaLeuSerGlyLeuGly---LysGluAla 502
Db 38438 TTGGTTTTTCTTAAGCTTGAATGCCCTGACAGATTTTCTGCTGTAATGAAAAATCAT 38379
QY 503 SerAspAlaGluValLeuAlaSerGluProValArgAlaTyrPheAlaAspTyrLeuLys 522
Db 38378 AGGCGAATAGATATCTTGCACAAACCCAAAGTACAGAGTGTTCGACAGTTTAAATT 38319
QY 523 ArgLeuAsnArgGluAlaThrGlyAsnAlaSerArgIleMetTyrValGlyLeuLeuAsp 542
Db 38318 CAGTACAAACAAAGATGCCAGGCGCATTCAAATCGSGTATCAATGTGATTAAATGACT 38259
QY 543 ThrProProSerIleAspArgLysGlyGluValThrAspArgGlySerIleAsnGlnArgAla 562
Db 38258 GAAGCACCAAGCTCCGATGACAGGTGAAGTAAACCATTAAGGCATTTAAATCAGAGTAAC 38199
QY 563 ValLeuGlnTyrArgSerAlaLysValAspAlaLeuTyrArgGlyGluAspGlnSerMet 582
Db 38198 ATCTTAAAGAGTGTGCTGACAGTGTGAAGAACTGTATTAACAACAAGATTAATCCA 38139
QY 583 Leu 583
Db 38138 TTG 38136

```

RESULT 11
CR543861_17/c
WPCOMMENT
Sequence split into 36 fragments LOCUS CR543861 Accession CR543861

Fragment Name	Begin	End
CR543861_00	1	110000
CR543861_01	100001	210000
CR543861_02	200001	310000
CR543861_03	300001	410000
CR543861_04	400001	510000
CR543861_05	500001	610000
CR543861_06	600001	710000

```

CR543861_07 700001 810000
CR543861_08 800001 910000
CR543861_09 900001 1010000
CR543861_10 100001 1110000
CR543861_11 110001 1210000
CR543861_12 120001 1310000
CR543861_13 130001 1410000
CR543861_14 140001 1510000
CR543861_15 150001 1610000
CR543861_16 160001 1710000
CR543861_17 170001 1810000
CR543861_18 180001 1910000
CR543861_19 190001 2010000
CR543861_20 200001 2110000
CR543861_21 210001 2210000
CR543861_22 220001 2310000
CR543861_23 230001 2410000
CR543861_24 240001 2510000
CR543861_25 250001 2610000
CR543861_26 260001 2710000
CR543861_27 270001 2810000
CR543861_28 280001 2910000
CR543861_29 290001 3010000
CR543861_30 300001 3110000
CR543861_31 310001 3210000
CR543861_32 320001 3310000
CR543861_33 330001 3410000
CR543861_34 340001 3510000
CR543861_35 350001 3598621

```

Continuation (18 of 36) of CR543861 from base 1700001 (CR543861 Acinetobacter sp. ADP1 cc

Alignment Scores:

Pred. No.:	6.8e-104	Length:	110000
Score:	1618.50	Matches:	307
Percent Similarity:	69.54%	Conservative:	97
Best Local Similarity:	52.84%	Mismatches:	174
Query Match:	52.51%	Indels:	3
DB:	1	Gaps:	3

US-09-750-986D-30 (1-589) x CR543861_17 (1-110000)

```

QY 5 GlnAlaLeuLeuProPheProGlyArgIleLeuGluArgLeuGlnHisTyrAlaLysThr 24
Db 33726 GAACCACTCAAGGCTTATCCGCAAAAAGTACCGATGCTTTGATTCATTTTGTAAAGTT 33667
QY 25 ArgProGlnGlnThrCysValAlaAlaArgAlaAlaAsnGlyLutPrArgArgIleSer 44
Db 33666 AAGCCCTAACATACCTTGGCCGCTTAAACGCAATACAAAGCAATGATCAGGCTCAGT 33607
QY 45 TyrAlaGluMetPheHisAsnValArgAlaIleAlaGlnSerLeuLeuProTyrGlyLeu 64
Db 33606 TATGCAAGAAACCTTGCAACGTGCTTGCGATTTTGACACAACTTTTGATCGAGCAACTTA 33547
QY 65 SerAlaGluArgProLeuLeuIleValSerGlyAsnAspLeuGlnHisLeuGlnLeuAla 84
Db 33546 AGCCAAAGACGACTTTGGTATTTATTCGGGAAAGATTTTGAACCTTAAGCTTTTCG 33487
QY 85 PheGlyAlaMetTyrAlaGlyIleProTyrCysProValSerProAlaTyrSerLeuLeu 104
Db 33486 ATGGGCGCGATGTTGGACAGGTGATACCATTTTGTGATTTTCGACAGCTCACTCAGTGTA 33427
QY 105 SerGlnAspLeuAlaLysLeuArgHisIleValGlyLeuLeuGlnProGlyLysLeuValPhe 124
Db 33426 TCTCAAGATTTTGGCAAAATTTAAACATGTGTTGATGTGCTCAACCCGAGGATGGTGTAT 33367
QY 125 AlaAlaAspAlaAlaProPheGlnArgAlaIleGluThrIleLeuProAspAspValPro 144
Db 33366 GCAAAATGATGACAAAGCTTTTGTAAAGCATTTCAAGCTTACGCAAGATCATATTGAA 33307
QY 145 AlaIlePheThrArgGlyLeuLeuAlaGlyArgArgThrValSerPheAspSerLeuLeu 164
Db 33306 ATTGTGACCAATTAAGGATTTATAGTGTGATCAAGTCTGTACATCGTTTCAGTCTTATTTA 33247

```

QY 165 GlnGlnProGlyGlyIleGluAlaAspAsnAlaPheAlaIleThrGlyProAspThrIle 184
Db 33246 GATACCCCT---GTGACCGATGTTCAAGAAcATTATGAAACGGTAGATGACATCAT 33190
QY 185 AlAllyPheLeuPheThrSerGlySerThrIleLeuProLysAlaValLProThrThrGln 204
Db 33189 GCCAAGTTCATATTCATCCGGTTCGACCAAGCTTCGAAAGCGTGTTCGACCAACAT 33130
QY 205 ArgMetLeuCysAlaAsnGlnGlnMetLeuGlnThrPheProValPheGlyGlu 224
Db 33129 TTTAGTCTGTATCATCAACAATCTCTCAAACTTCGAGATTTGGAAGACACA 33070
QY 225 ProProValIleValLeuPheProTyrPhePheThrPheGlySerHisAsnIle 244
Db 33069 CCGCTGTCTTACTGAGCTGTTGCTGTCATCATATTCGATGGAACCAATGTT 33010
QY 245 GlyIleValLeuIleValLeuGlyGlyIleThrTyrIleLeuAspArgLysProThrAlaGln 264
Db 33009 GGGATTCGGCTTATTAACGGCGGAAcATTATCATATCCGTAAGCCCGTCCGGA 32950
QY 265 GlyPheAlaGluThrLeuArgAsnLeuSerGluIleSerProThrAlaTyrLeuThrVal 284
Db 32949 AATTTGATGAAACGATTCGAAATTTAAAGAAATTTACCGACGGTTTCTTAAATGA 32890
QY 285 ProLysGlyTyrGluLeuValGlyAlaLeuGluArgAspSerThrLeuArgLysArg 304
Db 32889 CCCAAAGCGCTGGAGAAGAACTCACCGATGCACTTGAAAAAATGCCGAACCTCAAGATCGC 32830
QY 305 PhePheAlaArgMetLysLeuPhePhePheAlaAlaIleValLeuSerGlnGlyIleTyr 324
Db 32829 TTTCTTGGCAAGCTTAAATTTGTTTGTGCTGCTCATATCTGGAAGCAGGCTGG 32770
QY 325 AspArgLeuAspArgValAlaGluGlnHisCysGlyGlyLysArgIleArgMetAlaGly 344
Db 32769 AACCGATCGATTAATTAATGACACAGCAGCATTCGCGTAAATAATCCGATCATGAGTGGC 32710
QY 345 LeuGlyMetThrGluThrAlaProSerCysThrPheThrThrGlyProLeuSerMetAla 364
Db 32709 TTGGGATGAGCCAAACGGACACCATCGTGGCTTTTCCACAGGCCACCGGTAATGGCT 32650
QY 365 GlyTyrIleGlyLeuProAlaProGlyCysGluValIleValLeuValProValAspGlyLys 384
Db 32649 GGTTCATATTCGATCCCTGCTCCGGGCTGGAATCAATTAAGTCCCTATGTGTACAAA 32590
QY 385 LeuGluGlyArgPheHisGlyProHisValMetSerGlyTyrTyrParg--AlaProGlu 403
Db 32589 CTTGAAATTTGTGTGTCGTGCGCAAGCAGTCATGAAAGTACTGCGCCCTTAAGCGCAC 32530
QY 404 GlnAsnAlaGlnAlaPheAspGluGluGlyTyrTyrCysSerGlyAspAlaIleLysLeu 423
Db 32529 CACCAAAAGCAcAATTTTGTATGAAGACGGCTTTTCCATCTGCGCATGCGTACGCTG 32470
QY 424 AlaAspProAlaAspProGluLysGlyLeuMetPheAspGlyArgIleAlaGluAspPhe 443
Db 32469 GTTCGATCCAACACCTGACAGGGCTATGATGACACGGCGTATTCGAGAAAGATTTC 32410
QY 444 LysLeuSerSerGlyValIlePheValSerValGlyProLeuArgThrArgAlaValLeuGlu 463
Db 32409 AACTCAACACGGGTACCTTTGTGAATGTGGCAGCTTTAGCAACAAGCAGCATTCACAA 32350
QY 464 GlyGlySerTyrValLeuAspValIleValAlaAlaProAspArgLysCysLeuGlyLeu 483
Db 32349 GGTAAATTTGTGTAAGATGTTTGTATTAACGGTTCAAATCGAATGCGATTTGCGCTTC 32290
QY 484 LeuValPheProArgLeuLeuAspCysArgAlaLeuSerLysLeuGly--LysGluAla 502
Db 32289 TTGCTTTTCTTAAGCTGATGCTGCGAGATTTCGCTGCTGATATGAATAATCAT 32230
QY 503 SerAspAlaGluValIleValLeuSerGluProValArgAlaPhePheAlaAspTyrLeuLys 522
Db 32229 AGCGCAATAGATATCTTTCGAAACACCCTAAGTACAGAGTGTTCGACAGTTTAAAT 32170
QY 523 ArgLeuAsnArgLysAlaThrGlyAsnAlaSerArgIleMetTyrValGlyLeuLeuAsp 542

Db 32169 CAGTACAAACAAGATCCACGGCGACCTCAAACTCGGATTCAAATGTTGATTTAAAGACT 32110
QY 543 ThrProSerIleAspLysGlyValIleThrAspLysGlySerIleAsnIleArgAla 562
Db 32109 GAAGCACCAAGCTCATGCGATGAGTAAGCCGATTAAGGCAATTTAAATTCAGACTAAC 32050
QY 563 ValLeuGlnTyrPargSerAlaLysValAspAlaLeuTyrArgGlyLysArgIleSerMet 582
Db 32049 ATCTTAAACGTCGTCTGACGTCGCAAGAACTGTATTAACAACAACAGATTAATCCA 31990
QY 583 Leu 583
Db 31989 TTG 31987
RESULT 12
LOCUS BX572598 346879 bp DNA linear BCT 20-APR-2004
DEFINITION Rhodopseudomonas palustris CGA009 complete genome; segment 6/16.
ACCESSION BX572598 BX571963
VERSION BX572598.1 GI:39648490
KEYWORDS complete genomes.
SOURCE Rhodopseudomonas palustris CGA009
ORGANISM Rhodopseudomonas palustris CGA009
REFERENCE 1 Rhodopseudomonas palustris CGA009
AUTHORS Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Bradyrhizobiaceae; Rhodopseudomonas.
1 (bases 1 to 346879)
Larimer, F.W., Chain, P., Hauser, L., Lamerdin, J., Malfatti, S., Do, L.,
Land, M.L., Pelletier, D.A., Beatty, J.T., Lang, A.S., Tabita, F.R.,
Gibson, J.L., Hanson, T.E., Bobec, C., Torres, J.L., Peters, C.,
Harrison, F.H., Gibson, J. and Harwood, C.S.
Complete genome sequence of the metabolically versatile
photosynthetic bacterium Rhodopseudomonas palustris
Nat. Biotechnol. 22 (1), 55-61 (2004)
14704707
JOURNAL 2 (bases 1 to 346879)
PUBMED Larimer, F.W. and Harwood, C.S.
REFERENCE Rhodopseudomonas genome consortium
AUTHORS Direct Submission
CONSTRM Submitted (24-JUL-2003) Submitted on behalf of the Rhodopseudomonas
TITLE genome consortium, the DOE Joint Genome Institute, Production
JOURNAL Genomics Facility, 2800 Mitchell Drive, Walnut Creek, CA 94598,
USA, and the Genome Analysis Group, Oak Ridge National Laboratory,
1060 Commerce Park Drive, Oak Ridge, TN 37831, USA;
larimerf@ornl.gov
FEATURES
source Location/Qualifiers
1. 346879
/organism="Rhodopseudomonas palustris CGA009"
/mol_type="genomic DNA"
/strain="CGA009"
/db_xref="taxon:258594"
127..1005
/gene="dapA"
/locus_tag="RPA1571"
127..1005
CDS
127..1005
/locus_tag="RPA1571"
/EC_number="4.2.1.52"
/function="InterPro IPR002220 COGS COG3229"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="putative dihydrodipicolinate synthetase"
/protein_id="GI:39648491"
/db_xref="GI:39648491"
/translation="MMADPHGVPPYIVSPVDAEGRRADVNGRLCDLIQAGVHGILP
LSTSEFALVDAQREAVRATIEAQRVPVAVAGVASTVADVAQAATLEKLDG
ILALIEAYPLDQAQIESYFRAIDAEVPIVYINPQORSDLTLIDVLAIEHPRI
RVIKASTNTGSLSTINCGDALGVFSASAIH PAAVMLIGGVMAAGPACIARQSV
ALYEICKAQRMEALMLQKRLWRVNEAFKFLNACIKKAGLALQGVDDVPPIPPQAL
THBRKAVYKVLAEIA"
1230..3560
/locus_tag="RPA1572"

CDS
 1230. .3560
 /locus_tag="RPA1572"
 /function="InterPro IPR00674 COG6 COG1529"
 /note="observed by proteomics
 citation: Proteomics from Verberkmoes et al. (2003)
 unpublished"
 /codon_start=1
 /evidence=not experimental
 /transl_table=11
 /product="putative carbon-monoxide dehydrogenase large
 subunit"
 /protein_id="CAE27013.1"
 /db_xref="GI:39648492"
 /translation="WNILPGSMRFGAGQPVRLRELDQRLITGHGYLDKKAEGALMLV
 VMSRHADIVSIDTLEAKGKMGVEALVTADLADVGITPTLPIFEKPDGPMML
 PLRLALHEKFAVGEPAVAVAATQLAAQAMEIAVEYHELPAVDPPTALIDPAP
 AWAEHPNDIVAAMSGYGAADKADFAFAAHATVSLDVSORLISVASEPSTTAEIKK
 KTRGLILHQSGOTPGQTRDALAEALIKPKESIOYVSDICGGCGQCTGYPRDALVA
 YAAVLNKRKIKRKEGTDFIGTTHGRDLITSAFSDSKRVALATVTSIGGAYL
 AGAVIIPVLGPIFQVTGYVHLPVVHPIKAVMHTAVAGVYRAGAREAVYIYERLM
 DAARKLMDPRAIKRINVIKPSQLPYNTAAVGOVDSGAFHMDRAELSDMGFKA
 RKKAOKKGLLYRGVTSYIIEWTGGRAITENVKLHATLEGGVILSHSGOAMQGLLETT
 YTMIAALLETPLDKIDIVQCGNTDLIFGSGVSRSLPVGSTAAVAVSGEIMAAKREK
 ANTEIALEDIEDIEGGMULTIAGTDRKISLEPIAKKEGRTLSYDSGVEDSGEPWPK
 AHICEVDEPETSIVSRVRYTTVDVGNVAPMLVAGQIHGGVAGVGQALYEAALYN
 GDQLLTALSYGLDYDCIPRADNLPPIVNTLDPAPCKTPIFGAKGCGESGAIIGCPVCVH
 GVDLALAPGVTTLNTPLTPEKVMRAIDAKSAQA"
 3638. .4252
 /locus_tag="RPA1573"
 3638. .4252
 /locus_tag="RPA1573"
 /function="COG6 COG1704"
 /note="observed by proteomics
 citation: Proteomics from Verberkmoes et al. (2003)
 unpublished"
 /codon_start=1
 /evidence=not experimental
 /transl_table=11
 /product="Lema family"
 /protein_id="CAE27014.1"
 /db_xref="GI:39648493"
 /translation="MRKFLTYALALALSLNSCGYNTIQSDQSVKTSWESEVNOYOR
 RADLVILVNSVKGFAQQEDKVLGTVTAAKVAKGSIQATPEVLNDPAPAFQKFDQAGE
 LSSALSRLITVENYPOKSDLEFNLMAOEGEENRTVVARNRVIAVAVAIVTVRS
 VVTFTAMFGYGEKKNFTVENEKASITAPKVDFTPAPAPAK"
 3638. .3700
 /locus_tag="RPA1573"
 /note="Signal predicted by SignalP 2.0 HMM (Signal
 peptideprobability 0.998) with cleavage site probability
 0.396 atresidue 21"
 4257. .5180
 /locus_tag="RPA1574"
 4257. .5180
 /locus_tag="RPA1574"
 /function="COG6 COG1512"
 /codon_start=1
 /evidence=not experimental
 /transl_table=11
 /product="DUF477"
 /protein_id="CAE27015.1"
 /db_xref="GI:39648494"
 /translation="MTATIPRAVRAGKATLALLCVALLALAPFARADVAVPYLTCRVD
 OTGTASADVAVLDDKLKAFEPARKGSOIAVLIVPTTPEATIEORSIRVAEMKIGRKK
 IDDGAILLVAKRKRLRIEYVCGLEGALPDYTKARIIDEITTPKFGCDPDGCEAGV
 DMKTIIVGVAEPLPTPOQHEHMAPEASMDLTGMALPILGTVLNGFLKALGWTGGS
 GLTIGTIVGVAEPLFGLVWYIALVAVGFAFLALATVILFGGPTISSGRSGGGGTGGS
 SSGSSWSGSSSDSGSGGGGSGFGGAGAGS"
 4257. .4352
 /locus_tag="RPA1574"
 /note="Signal predicted by SignalP 2.0 HMM (Signal
 peptideprobability 1.000) with cleavage site probability
 0.947 atresidue 33"
 order(42930. .4358,4824. .4880,4899. .4952,4962. .5030)

Alignment Scores:

```

/locus_tag="RPA1574"
/note="T4 probable transmembrane helices predicted
byTMHMM2.0"
5194. .5691
/locus_tag="RPA1575"
5194. .5691
/locus_tag="RPA1575"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="conserved hypothetical protein"
/db_xref="GI:39648495"
/translacion="MGIRRIQKHLFTSRGAVHRAFSPPADFALERAIKASETHVGOI
RFVEGALDGAFLRPDQPVVERADITISOLRIWPTENNNGVLITLADRNVETIAD
GIDRVGADTWEAICREMEAEFRREGQGLRGIDITRIALAEHFPSSGSPNEIPD
APVVV"
complement(5773. .6384)
/locus_tag="RPA1576"
complement(5773. .6384)
/locus_tag="RPA1576"
/function="InterPro IPR004045:IPR004046 COG6 COG625"
/note="observed by proteomics
citation: Proteomics from Verberkmoes et al. (2003)
unpublished"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative glutathione S-transferase"
/protein_id="CAE27017.1"
/db_xref="GI:39648496"
/translacion="MKLYDSIGPNRLVRMEIAEKGLTPTQTLDIRAGENRKEPHLA
RNPQGTPTLELDNGSEFLAEVTAICEVLEEKHPQPLIGTTPBEARACMWTRRIDLY
ICEBLTNGPFAEGLKMEODRVCVPEAPGLKKIADRIAMNEOKKKEVYCGARLL
TLAALLXGMLDFGAQVQGPLAPANTTIAEMFKRIARPAKA"
6593. .7564
/locus_tag="RPA1577"
6593. .7564
/locus_tag="RPA1577"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="CAE27018.1"
/db_xref="GI:39648497"
/translacion="MDENRNPTEIGEOQTGAATDGAONNAGTPBASASAPDGTTPAN
TPNAGGAKPDEAPKPKSGSEIKSDRYSRPRPGVTITMARSPPEDLYRHMDDELVEVRA
PDEASQAVHREFGVKCLAAVAATLTALAGAVGALATLSLRSDAETTPASAAQVKA
TLDQVQALQIETLSALKADLDAAQVSTADQGLADITSEKLEKQASPAVATIALSTSE
VEKTRIAOVOPFAPAPAPANAAVAPAKDVTGSIPTPKAANAAGRLPTVEGWLRTDVG
GAVVEGRSGTVEYVAGDPPIGVGRVDAIRQDGRWVVTSGRLIVAR"
6947. .7015
/locus_tag="RPA1577"
/note="T1 probable transmembrane helix predicted
byTMHMM2.0"
7593. .7621
/note="TransTerm prediction, 98+%"
complement(7721. .8494)
/gene="fpr"
/locus_tag="RPA1578"
complement(7721. .8494)
/gene="fpr"
/locus_tag="RPA1578"
/EC_number="1.18.1.2"
/function="InterPro IPR001433:IPR001834 COG6 COG1018"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="ferridoxin-NADP+ reductase"
/protein_id="CAE27019.1"
/db_xref="GI:39648498"

```

Pred. No.:	1 04e-91	length:	346879
Score:	1457.00	Matches:	222
Percent Similarity:	65.40%	Conservative:	86
Best Local Similarity:	50.52%	Mismatches:	14
Query Match:	47.27%	Indels:	6
DB:	1	Gaps:	2

US-09-750-986D-30 (1-589) X BK5/2598 (1-346879)

Oy	1	MetAspSerLeuGluAlaLeuLeuProPheProGlyArgIleLeuGluArgLeuGluHis	20
		:::	:::
Db	256341	GTGCCTCGACCAACACAGCTTACCCAGCTATCCGGTCCGATTACCGACGGCTGCATCAC	256400
Oy	21	TrpAlaLysThrArgProGluGlnThrCysValAlaAlaArgAlaAlaLeuGlyIleTrp	40
		:::	:::
Db	256401	TTGCCGAGACGGCGCCCGACGGGGTTCATGGCCGGAACGGTAAAGGGCGGCTGG	256460
Oy	41	ArgArgIleSerTyrAlaGluMetPheHisAsnValArgAlaIleAlaGlnSerLeuLeu	60
		:::	:::
Db	256461	CGACAGATCACTATGACCGAGATGCTGCAGCCGCGCAACATCCCTCGGGCTGATATC	256520
Oy	61	ProTyrGlyLeuSerAlaGluArgProLeuLeuIleValSerGlyAsnAspLeuGluHis	80
		:::	:::
Db	256521	GCGCGCGGACTCTCGGCCGAACGGCCGGGATGATCTCTCCGGTAAATTCATCGACCAAT	256580
Oy	81	LeuGlnLeuAlaPheGlyAlaMetTyrAlaGlyIleProTyrCysProValSerProAla	100
		:::	:::
Db	256581	GCGATGCTGATGTTCTGGCGCGCTGATGGGGCGGCGCATGTCGGGTGTGCGCCGG	256640
Oy	101	TyrSerLeuLeuSerGlnAspLeuAlaLysLeuArgHisIleValGlyLeuGluGlnPro	120
		:::	:::
Db	256641	TATTCGCTGCTGTCCAGGATTACGGCAAGCTGCGCCATATGTCGGCTGCTGACTCCG	256700
Oy	121	GlyLeuValPheAlaAlaAspAlaAlaProPheGlnArgAlaIleGluThrIleLeuPro	140
		:::	:::
Db	256701	GGCGTGAATCTTCCCGGATCACTACACTGCTTCCGCCGCCGATCTCCGCCACCGTGGC	256760
Oy	141	AspAspValProAlaIlePheThrArgGlyIleLeuAlaGlyArgArgThrValSerPhe	160
		:::	:::
Db	256761	GAGGATGTCGAGCTCGCTGCCACGCGCGCGAGTGAAAGGGCGCAAGCTGACGTGCTT	256820
Oy	161	AspSerLeuLeuGluGlnProGlyGlyIleGluAlaAspAsnAlaPheAlaIleThrGly	180
		:::	:::
Db	256821	GCCGAATCGCTGGCGACGCGCGGAACATCCGAACTCCGCCGCGCAAGCAGAGGATCGC	256880
Oy	181	ProAspThrIleAlaLysPheLeuPheThrSerGlySerThrIlyLeuProLysAlaVal	200
		:::	:::
Db	256881	CACGACACCATGCTTAAGTTTTCGTCGACGTGGGATTCACCGGCAATCCGAAGCGGTG	256940
Oy	201	ProThrThrGlnArgMetLeuCysAlaAsnGlnMetLeuLeuGlnThrPheProVal	220
		:::	:::
Db	256941	ATCATATACGACCGGATGATCTGCGCCATCAAGTATGATCCGAGGAGATGCGGCTTC	257000
Oy	221	PheGlyGluGlnProProValLeuValAspTrpLeuProTrpAsnHisThrPheGlyGly	240
		:::	:::
Db	257001	CTGAAGAAGACCCCGGTGATCTGCGACTGCTGCCGTGATATACACTTCCTGCGCGC	257060
Oy	241	SerHisAsnIleGlyIleValLeuTyrAsnGlyGlyThrTyrIleLeuAspAspGlyLys	260
		:::	:::
Db	257061	AACCAATATGCTGTGACGCTGTTCACCGCGGCTCATGTATATTCACACGCGCAAG	257120
Oy	261	ProThrAlaGlnGlyPheAlaGluThrLeuArgAsnLeuSerGluIleSerProThrAla	280
		:::	:::
Db	257121	CCGACGCGCGGGGTGCTTCACCACTCCGGAATCTGGGAGATCCGGCGGACGGTGG	257180
Oy	281	TyrLeuThrValProLysGlyIleTrpGluIleLeuValGlyAlaLeuGluArgAspSerThr	300
		:::	:::
Db	257181	TATTTCAACGTTCCGAAGGGCTACAGATGCTGCTGCCGTGTGCGCAAGACCGACAG	257240
Oy	301	LeuArgGluArgPhePheAlaArgMetLysLeuPhePhePheAlaIleAlaGlyLeuSer	320
		:::	:::
Db	257241	TTGCCAATATGCTTCAACCGCGCTGCATGCGATCTTCTTCCGCGCGCAAGCTTCGG	257300

Oy	321	GLHGLYLITPRAAPRPRLEUASPRPRVALAAGLUGLNHLSYSGYLGLUALRGLPARG	340
Db	257301	GGCGATGTCGGAACGGGCTTACACAGAGTCGGCGGTCCGGAGACCGCGCGGGGTCCG	257366
Oy	341	MetMetAlaGlyLeuGlyYmerThiGluThrAlaProSerCysrhrPheThrThiGlyPro	360
Db	257361	ATGCTCACCGGCGCTTGCGCCACCGAGACCGGCGCTTCTCATGTGCGTACCCGACAG	257420
Oy	361	LeuSerMetAlaGlyYrTrlLeuGlyLeuProAlaProGlyCysGluValLyLeuValPro	380
Db	257421	ACCAAGTCGCTCCGGCCATGTCGGCTGCGCGGTCCGGCCCAACGAGGCCAACCTGTGTCG	257480
Oy	381	ValAspGlyLyLeuGluGlyLeuArgPheHlaGlyProHISValMetSerGlyYrTrlPArg	400
Db	257481	AACACAGGCAAGCTGCAGATGCCGCCCAAGGGGCCAAACATCACCCCGCTATTGGCGC	257540
Oy	401	AlaProGluGluHnaHnaAlaGlnAlaPheAspGluGluGlyYrTrlCysSerGlyAspAla	420
Db	257541	GGCCCCGAGCTGACCGATAGGCGGTTCGACGAGGAGGCTTCTACAGCTCAACGATGCG	257600
Oy	421	ILeLyLeuAlaAspProAlaAspProGluGlyLeuMetCpHeAspGlyArgTrlLea	440
Db	257601	CTGAACCCGGTCGATGCCAACACACTTTTCGCGCGCTTCGATTCGACGCGCGGATCTCG	257660
Oy	441	GluAspPheLyLeuSerSerGlyValPheValSerValGlyProLeuArgThrArgAla	460
Db	257661	GAAGACTCTCAAACTGGCGGTGGGACCTGGGTGACGCGGTGGCGCTACGCGCCAACTTC	257720
Oy	461	ValLeuGluGlyYrGlySerYrValLeuAspValValAlaAlaProAspArgGluCys	480
Db	257721	ATTGCGCGCGCGCGCTCCCTCGGTGGCGACGTCGTGATCGCGGGCTGACCGGAGTTAC	257780
Oy	481	LeuGlyLeuLeuValPheProArgLeuLeuAsp-----CysArgAlaLeuSerGly	497
Db	257781	GTCACCGCGCTGGCG-----ATCTCGATCCCGACCGGCTGCAAGCTGATCAATGCG	257831
Oy	498	LeuGlyLyLeuGluAlaSerSerPheGluValLeuAlaSerGluProValArgAlaTrpPhe	517
Db	257832	ACGCTGCGCGCTGGAGACCTGCGCGGACATGCGCGCGCGACCATCTGATCGGAGGCGTTT	257891
Oy	518	AlaAspTrpLeuLySArgLeuHnaArgGluAlaThrGlyAsnAlaSerArgTrlMetTrp	537
Db	257892	CGGAGCGCTTCCGCCACGCTGCTGACGACGAGCGGCGCTGCTCAACCGCGTACCGCG	257955
Oy	538	ValGlyLeuLeuAspThrProProSerTrlLeuAspLySGlyGluValThrAspLyGlySer	557
Db	257952	GCGGTGCTGCTGGCGAACCGCTGTCGATCGACAGAGGGGAGATCACCGAACAGGCGCTCG	258011
Oy	558	ILeAsnGlnArgAlaValLeuGlnTrlPArgSerTrlAlaLyValAspAlaLeuTrlYr	575
Db	258012	GTCAACCAAGCGGCGCTGCTGGAAATACGCGGCTGTTGATCGCGGATCTTTAC	258065
RESULT 13			
LOCUS	AP005963	299950 bp	DNA linear BCT 12-SEP-2003
DEFINITION	Bradyrhizobium japonicum USDA 110		complete genome, section 29/31.
ACCESSION	AP005963	BA000040	
VERSION	AP005963.1	GI:27355938	
KEYWORDS			
SOURCE	Bradyrhizobium japonicum USDA 110		
ORGANISM	Bradyrhizobium japonicum USDA 110		
REFERENCE			
AUTHORS	1 Kaneke, T., Nakamura, Y., Sato, S., Minamiawa, K., Uchiumi, T., Sasamoto, S., Watanabe, A., Ideesawa, K., Itiguchi, M., Kawashima, K., Kohara, M., Matsumoto, M., Shimizu, S., Tsunokawa, H., Wada, T., Yamada, M. and Tabata, S.		
TITLE	Complete genomic sequence of nitrogen-fixing symbiotic bacterium Bradyrhizobium japonicum USDA110		
JOURNAL	DNA Res. 9 (6), 189-197 (2002)		
MEDLINE	22484998		

PUBMED REFERENCE	12597275
AUTHORS	Kaneko,T., Nakamura,Y., Sato,S., Minamisawa,K., Uchiumi,T., Sasamoto,S., Watanabe,A., Idesawa,K., Iriyuchi,M., Kawashima,K., Kohara,M., Matsumoto,M., Shimpō,S., Tsuruoka,H., Mada,T., Yamada,M. and Tabata,S.
TITLE	Complete genomic sequence of nitrogen-fixing symbiotic bacterium Bradyrhizobium japonicum USDA110 (supplement)
JOURNAL	DNA Res. 9 (6), 225-256 (2002)
MEDLINE	22485002
PUBMED	12597279
REFERENCE	3 (bases 1 to 299950)
AUTHORS	Kaneko,T.
TITLE	Direct Submission
JOURNAL	Submitted (20-NOV-2002) Takakazu Kaneko, Kazusa DNA Research Institute, The First Laboratory for Plant Gene Research; 2-6-7 Kasuga-Kametarai, Kisarazu, Chiba 292-0812, Japan (E-mail:kaneko@kazusa.or.jp URL:http://www.kazusa.or.jp/rhizobase/, Tel:81-438-52-3935(ex.2338), Fax:81-438-52-3934) Location/Qualifiers
FEATURES	1..299950
source	/organism="Bradyrhizobium japonicum USDA 110"
	/mol_type="genomic DNA"
	/strain="USDA110"
	/db_xref="taxon:224911"
CDS	/db_xref="bar7655"
gene	36..278
	/gene="bsr7655"
	/note="ORF ID:bsr7655
	hypothetical protein"
	/codon_start=1
	/transl_table=11
	/protein_id="BACS2920.1"
	/db_xref="GI:27355939"
	/translation="MMQATVDSDSAIDEMAAVYDRGRGLDVVNNAAGHEGAPAA
	SITDKKRWSTVDGVGYGCRAALPTLRPGARS"
gene	275..637
	/gene="bl1r7656"
	275..637
CDS	/gene="bl1r7656"
	/note="ORF ID:bl1r7656
	hypothetical protein"
	/codon_start=1
	/transl_table=11
	/protein_id="BACS2921.1"
	/db_xref="GI:27355940"
	/translation="MNRVSYSGTGDGMGMSPYNAAKGAVVNLTRALADLGGKIRAV
	AVCPSLTFRGTEDMDDEKLAKFAERIPLGVCPEBEVAIVIAFLSEDSAFMTGA
	NNAVGGVASANGCOPQE"
	complement(660..1346)
	/gene="bl17657"
	complement(660..1346)
gene	/gene="bl17657"
	/note="ORF ID:bl17657
	putative phosphoglycolate phosphatase"
	/codon_start=1
	/transl_table=11
	/protein_id="BACS2922.1"
	/db_xref="GI:27355941"
	/translation="MMDMEDSELPKAIFDLDTLLSDVDLHALAQEAWEFGHDVT
	FEQVRISIGKGKLIPLVLSDQREHGHLIEWGNRFSKTYLPRLRPFAVSOL
	ORVREGLRTAVASAKDELKKYLIDAIISDLYVATSSEDVDSKPDI FEVYLK
	KLGKSGEAYVIGDTPYDAIARAKGVSALIGVCGFTENSIREACGEVYVGPOTLY
	ARFDYDALTK"
	complement(1503..1880)
gene	/gene="b117658"
	complement(1503..1880)
CDS	/gene="b117658"
	/note="ORF ID:b117658"
	/codon_start=1
	/transl_table=11
	/product="two-component response regulator"
	/protein_id="BACS2923.1"
	/db_xref="GI:27355942"
	/translation="NSNALRVLVVEDEALICSFIEDALSOGCFACSVHSGEAMSTFF
	RGRGECRALTLRDVNLGGISGEWELARQIREITGFVPVWTTSAPDWKSQVDSL
	LIEKPPAPQLAAVSQLDTRA"
	complement(1987..2319)
gene	/gene="b117659"
	complement(1987..2319)
CDS	/gene="b117659"
	/note="ORF ID:b117659
	hypothetical protein"
	/codon_start=1
	/transl_table=11
	/protein_id="BACS2924.1"
	/db_xref="GI:27355943"
	/translation="MAMDREFFSLIGSDNEGTVNVYGADKEKYIERWIDKVSCK
	VSHAVLSFPDGLRI GDVHYPLPWQALKYDTNLGCVYTGITQDHILRSAPKADESMMW
	SDPAPRAP"
	complement(2397..2825)
gene	/gene="b117660"
	complement(2397..2825)
CDS	/gene="b117660"
	/note="ORF ID:b117660
	unknown protein"
	/codon_start=1
	/transl_table=11
	/protein_id="BACS2925.1"
	/db_xref="GI:27355944"
	/translation="MRBRPMKPCRIITDTTIIVPNRLQGAAAAPRCAGTLDLGR
	ACRGSYAIVRSITYQGTGEGIPESRILRGRRSPVSDPVVDIAMKRFGHEHELHQ
	PGHDLVHERRADPGEARQLPQGFALQGFHARTPLIQ"
	complement(2979..3422)
gene	/gene="b117661"
	complement(2979..3422)
CDS	/gene="b117661"
	/note="ORF ID:b117661
	hypothetical protein"
	/codon_start=1
	/transl_table=11
	/protein_id="BACS2926.1"
	/db_xref="GI:27355945"
	/translation="MLMGCEFAENYFI LDDMGKRVSLLRPMKDFPYDESGRTVSL
	VRLLRIRQRDIRQGTYPFPNDMDRYQQGILIFAHDAQTLLATNFGDITRTV
	PFWFPRKGVNVEELGVALLDNVPALQOVLDPISHYRWLTJ"
	3427..3654
gene	/gene="bar7662"
	3427..3654
CDS	/gene="bar7662"
	/note="ORF ID:bar7662
	unknown protein"
	/codon_start=1
	/transl_table=11
	/protein_id="BACS2927.1"
	/db_xref="GI:27355946"
	/translation="MPLEARSGISAKUGSRMHLCGRYYGGRISVHGKEAVHVANKGQ
	SHGRYPFDPARGPSAKFRSSGSSPGDLLQRDW"
	complement(4162..4824)
gene	/gene="b117663"
	complement(4162..4824)
CDS	/gene="b117663"
	/note="ORF ID:b117663
	unknown protein"
	/codon_start=1
	/transl_table=11
	/protein_id="BACS2928.1"
	/db_xref="GI:27355947"
	/translation="MKRMRYASISTVIYAVGLVCAAGLVWGQSHVPDPAIASVTRAP
	ELMERAMHLPVATPFOQRVDMQSNGSRCGPAVANAYRSLCAARTSGKTACTGSCM
	TGVCIWGLTDIELAKVAQNTSRKTVLRDLSENQFLELRANDPGRRTYVINDRARR
	IRGASGHHSPIGCVFEAEDLVFVDVFNQPMLVERKRLFDAVNTLDDOKRGLLL
	IE"
gene	complement(7319..9061)

CDS
 /gene="b117664"
 complement (7319, .9061)
 /gene="b117664"
 /note="ORF ID: b117664"
 putative adenylate cyclase 3 (EC 4.6.1.1)"
 /codon_start=1
 /transl_table=11
 /protein_id="BAC52929.1"
 /db_xref="GI:27355948"
 /translation="MTEQRYQRLAIIADVGVYSALMORAREATYAFERLRELI
 EPLSHDGLRTKTDGALAFRSPPLAARCATEMODHAEBSGPKLRIGNLGV
 IVQDGDLYDGINIAVRLEGIDPQGITLSEKVSVESEKLVGEDEGEQJKNIA
 KPVRAVAVRGASSALTERLSAARPLPKSLIAPLPENNSGDPDEQFVADKEIT
 TALSRKMLFVIARNSFTYKGAIVNVKQGHILGVYILIEGSRKKAQVIRAGQI
 DAITGTHIWDREEDLTDFALODVVALVAVQRMQTEIATARRPENLITAY
 DFRSMQOQYVLTRESLEAETKLARALIDRPSGAALADCHASNVALGVANP
 EFPKAKIARLARIASLDGDPDPTLSRVAVSAQVSDSPAEIEMADRAVALPNMR
 TNRNKGVTYMAGLPEBAIRSFERRAIRMSFLDPLMTFTIGTGHVGLRRPDAITV
 LKRALQNPSFMPACRGLASFAHLGRDAHAHAARVLEIDPTTISGHIAVRPKSA
 KMFIEGLRKALPE"
 9440..10444
 /gene="b1r7665"
 9440..10444
 /gene="b1r7665"
 /note="ORF ID: b1r7665"
 putative methyltransferase (EC 2.1.1.-)"
 /codon_start=1
 /transl_table=11
 /protein_id="BAC52930.1"
 /db_xref="GI:27355949"
 /translation="MTSNPWLTLLGLINGFOVQAIVHASTLIADHLKDSHSAEL
 APLTQSHPALYRLRLNALPAVGFEEDEDFALTMDGLTRDSATPIGAMAEVGS
 PYWSTMGHLLIRITGENAFOSLNGDKILNRYAERPEEAVFDRAMTESRGAEDV
 IGAVDSAFHHVVVDVGGSGRLMAAILTLHARMOGTLPDQGVLAGAEVLEAHGVND
 RCMVSGSPFESVPBGSDAYVMKAVIHDMDDDAIATLKACRAMETATCVLIERII

Alignment Scores:

Pred. No.: 8, 6e-88 Length: 299950
 Score: 1401.00 Matches: 293
 Percent Similarity: 64.53% Conservative: 80
 Best Local Similarity: 50.69% Mismatches: 199
 Query Match: 45.46% Indels: 6
 Gaps: 4

US-09-750-986D-30 (1-589) x AP005963 (1-299950)

QY 1 MetArgSerLeuGluAlaLeuLeuProPheProGlyArgGileuGluAlaGluLeuHis 20
 Db 199280 CTGCGCGCGAAGCAGCGCGCTCGGAGCTATTCGTCGCGATCACCGACCGTTCGATCAT 199339
 QY 21 TTPAlaLeuThrArgProGluGluInThrCysValAlaAlaAlaArgAlaAlaAenGlyTyr 40
 Db 199340 TGGGCGACGACGACCGCGACCGCGTTCATGCGGAGCGCGAAGCGCGCGCGCTGG 199399
 QY 41 ArgArgIleSerTyrAlaGluMetPheHisAsnValArgAlaIleAlaInserLeuLeu 60
 Db 199400 CGCAAGATCACTATGCGCGAGCTGCTCACCGACCGCGACATCGCTCGGCGCTATC 199459
 QY 61 ProTyrGlyLeuSerAlaGluArgProLeuLeuIleValSerGlyAsnAspLeuGluHis 80
 Db 199460 CACCGCGCGCTTTCGCGGACCGCGCGGTGATCTCTCCGCGCAATTCGATCGACCAT 199519
 QY 81 LeuGlnLeuAlaPheGlyAlaMetTyrAlaGlyIleProTyrCysProValSerProAla 100
 Db 199520 GCGCTTTGGCTTTCGCGCGGCTTCTATGCGCGGCTTCGTTGCGCCGCGGTGCGCGCC 199579
 QY 101 TyrSerLeuLeuSerGlnAspLeuAlaIleLeuArgHisIleValGlyLeuLeuGlnPro 120
 Db 199580 TATTCGCTGGTGTGCAAGATTCACGCAAGCTCTCCATCTGATGAAAGCTTTCGACGCC 199639
 QY 121 GlyLeuValIleAlaAspAlaAlaProPheGlnIleArgAlaIle-----GluThrIle 138
 Db 199640 GGCCTCGTCTTCGCGCGAGCGCGACAGATTCGCGATGCGCTCGTCCCAATGTTTCG 199699

QY 139 LeuProAspAspValProAlaIlePheThrArgGlyGluLeuAlaGlyArgArgThrVal 158
 Db 199700 CTGCGGACCGAGATTCGCGCTCTAC-----GGCAATGTGCGCGCGCGACGTACG 199753
 QY 159 SerPheAspSerLeuLeuGluInProGlyGlyIleGluAlaAspAsnAlaPheAlaAla 178
 Db 199754 CTGCTCGCGGACCTTCAGGACCGCGATCCGCGGAGCTCTCCACAAAGTCCACGGCAG 199813
 QY 179 ThrGlyProAspThrIleAlaIlePheLeuPheThrSerGlySerThrIlePheProLys 198
 Db 199814 ATGCGCCGAGATTCGATTCGGAATTCCTGCTGACCTCGGCTCGACCGGCAATCCCAAG 199873
 QY 199 AlaValProThrThrGlnArgMetLeuCysAlaAsnGlnIleMetLeuLeuGlnThrPhe 218
 Db 199874 GCGCATATCAACCCAGCGGATGATCTGGCCCAACAGGTATGCTGCGCGAGACCTTC 199933
 QY 219 ProValIleGlyGluLeuProProValIleValAspTyrLeuProTyrPheHisIleThrPhe 238
 Db 199934 GCCTTCCTCAAGGACGAGCGCGCTGCATCATGCACTGGCTGCGGAAATCACACCTTT 199993
 QY 239 GlyGlySerHisAsnIleGlyIleValLeuTyrAsnGlyGlyThrTyrTyrLeuAspArg 258
 Db 199994 GCGCGAACCACATATCGGCTGACCGCTTCACATGGGAGCTTCGATGTTCTGGACGCC 200053
 QY 259 GlyLysProThrAlaGlnGlyPheAlaGluThrLeuArgAsnLeuSerGluIleSerPro 278
 Db 200054 GGCACCGCATGCGCGGTGGCAGAGAGACCGTGGCATTCACAGAGATTCGCGCG 200113
 QY 279 ThrAlaTyrLeuThrValProLysGlyTyrGlnGluLeuValGlyAlaLeuGluAlaGly 298
 Db 200114 ACCGTATTTTCAACCTGCGGAAAGGCTATGATCCCTGCTGCTTATGCGCGACGAC 200173
 QY 299 SerThrLeuArgGluArgPhePheAlaArgMetLysLeuPhePheAlaAlaIleGly 318
 Db 200174 CAGGCTCGGCGCGAATTCCTGACCGGCTGCATGCGATGTTCTTCTGCGCGCGCG 200233
 QY 319 LeuSerGlnGlyIleTyrAspArgLeuAspArgValAlaGluGlnHisCysGlyGluArg 338
 Db 200234 CTGTCCCGTGTGCTGCGGACGAGCTCGACAGCTCGCGGTGAAGAAAAGCTATCGC 200293
 QY 339 IleArgMetMetAlaGlyLeuGlyMetThrGluThrAlaProSerCysThrPheThrThr 358
 Db 200294 GTGCGATGCTGACCGCGCTTGGCGGACCGGACCGCGCTTCTTCATGTCGCTCAAT 200353
 QY 359 GlyProLeuSerMetAlaGlyTyrIleGlyLeuProAlaProGlyCysGluValLysLeu 378
 Db 200354 CCGCGACGACGCGCTCGGACGATGTCGCTCCGCTCCGCGAAGCGCGAAGCTG 200413
 QY 379 ValProValAspGlyLysLeuGluGlyArgPheHisGlyProHisValIleMetSerGlyTyr 398
 Db 200414 GTCCCGAACAACGCAAGCTCGAGGTCGCGCGCAAGGGCGCAACGTTGATCGCGGCTAC 200473
 QY 399 TTPArgAlaProGluGlnAsnAlaGlnAlaPheAspGluGluGlyTyrTyrCysSerGly 418
 Db 200474 TGGCGCGAGCGGATTCACCGCGAATATCTTCGAGGAAGAGATTCACAAAGTTGGGC 200533
 QY 419 AspAlaIleLysPheLeuAlaAspProAlaAspProGlnIleGlyLeuMetPheAspIleArg 438
 Db 200534 GACGCGCTCAAGGCTCCGATCCGATGATCTCAACCGCGGCTTCGATTCGACGGCGCG 200593
 QY 439 IleAlaGluAspPheLysLeuSerSerGlyValIleValSerValGlyProLeuAlaGlyThr 458
 Db 200594 GTACGAGAGACTTCACAGCTCGGACCGGATCACTGGGTCAAGTGTGCGCCCTTGGCGCGG 200653
 QY 459 ArgAlaValLeuGluGlyLysSerTyrValIleAspValValAlaAlaAspArgArg 478
 Db 200654 CGCTGTGTCGCGCTCGCGACCGCTGTGCGGAGAGTGTGTATCCGCGCGCATCAACCGC 200713
 QY 479 GluCysLeuGlyLeuLeuValIlePheProArgLeuLeuAspCysArgAlaLeuSerGlyLeu 498
 Db 200714 GACGAGGTCTCGCGGTGTGTGTGTCGACCTTCGACGAGGCTGTAGGTGTCAACCGGACG 200773

Oy	499	glytysglu1aseraapalagluvalleum1asergupro---ValArgAlatrphc	517
Db	200774	---CTGCCGCGCGACACTCTGCTCGCGACGCCGCGCGTGTGGAGGCTTT	200830
Oy	518	AlahepTleu1yAArgLeuanaArg1u1aThrg1yAsna1aseArg11emetrp	537
Db	200811	CGCGAGCCCTGACCCCTTCTCTCGCGCTCGGCAACCGGTTCTTCACCCGATCAGCGGC	200890
Oy	538	Valgl1yLeu1euaepThrProProSer11eaplyg1ygluval1ThraeplyeGlySer	557
Db	200891	GCATCTCGATGACACCCGCTCTCGATCGACAAGGCGAGGTCAACGACAGAGGCTCG	200950
Oy	558	11eang1na1ga1aValleuGln1trpArgser1a1yVala1a1a1e1uTy	575
Db	200951	ATCAACGAGCGCGCTGAGCATCGCGCGGCTGATCGACGACTCTAC	201004
RESULT 14			
LOCUS	AE008067/c	10173 bp	DNA linear BCT 28-MAY-2004
DEFINITION	Agrobacterium tumefaciens str. C58 circular chromosome, section 125 of 254 of the complete sequence.		
ACCESSION	AE008067		AE007869
VERSION	AE008067.1		GI:15156484
KEYWORDS			
SOURCE	Agrobacterium tumefaciens str. C58		
ORGANISM	Agrobacterium tumefaciens str. C58		
REFERENCE	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.		
AUTHORS	1 (bases 1 to 10173)		
TITLE	Hinkle,G., Slater,S.C. and Goodner,B.		
	Complete Genome Sequence of Agrobacterium tumefaciens C58 (Rhizobium radiobacter C58), the Causative Agent of Crown Gall Disease in Plants		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 10173)		
AUTHORS	Hinkle,G., Slater,S.C. and Goodner,B.		
TITLE	Direct Submission		
JOURNAL	Submitted (14-AUG-2001) Cereon Genomics, 45 Sidney Street, Cambridge, MA 02139, USA		
FEATURES	Location/Qualifiers		
source	1..10173		
	/organism="Agrobacterium tumefaciens str. C58"		
	/mol_type="genomic DNA"		
	/strain="C58"		
	/isolate="Cereon"		
gene	/db_xref="taxon:176299"		
	complement(100..1263)		
CDS	/gene="AGR_C_2609"		
	complement(100..1263)		
	/note="branched-chain amino acid ABC transporter, periplasmic amino acid-binding protein - Deinococcus radiodurans (strain R1)"		
	/codon_start=1		
	/transl_table=11		
	/product="AGR_C_2609p"		
	/protein_id="AAK87205.1"		
	/db_xref="GI:15156485"		
	/translation="MGKTKMKRITNMTATVALTTTGAAPADVTIGTIPPTGPGA		
	ALGIPLKQISIFWPDIDGKTKYLLDADGPPSVATTNARRFASDDKVDVIVSALT		
	PAALAVAGVANETEVPFALSP1PPTATKKNFVMPQSTVLVSKMFAMQKQDVKRV		
	AFIGFADYVGOOWDALKATKDYGLEVAERARADTSVGOVLKSLARPAIFI		
	GASGTGALPOIGVERGFKGVYVSGATFDFLR1AGKAVDITVFSGVMAEDQ		
	PDSAPTKOGLAYLKAEFGKPNRTFGSHWMDVWVILKKAVALKKKPKPAEF		
	RAARDAETSGPNATHGYEFTADNHYTGDKGVULLTAADKFLVLPNK"		
	complement(1317..2960)		
gene	/gene="AGR_C_2611"		
	complement(1317..2960)		
CDS	/gene="AGR_C_2611p"		
	/note="probable acyl-CoA dehydrogenase - Deinococcus radiodurans (strain R1)"		
	/codon_start=1		
	/transl_table=11		
	/product="AGR_C_2611p"		
	/protein_id="AAK87206.1"		
	/db_xref="GI:15156486"		
	/translation="MKSSGMWARSACGRDVMGSPSTQATMTLTKATSGMRR1QSLR		
	GDCCDALVVA11LNEARFAFVLEPLMTPADRGCTVADRVVPTGFADAFRRHAD		
	GMLADVPEREGQALPLVQOACAPLPERGCVLAAMSGSTRBAACH1LAEETGAVAL		
	DEWVPRLAAGEMAATTCISRPDAGSDALRRSAVPEEGMARI1EGKTIW1SFGDHMA		
	MR1GCHLLARTGETRTGRGSLFLVPEIDGNGVTVERT1EERKMGHGSPCLARPD		
	GANGLL1GEEGRGLAQ1PAM1ELMRLOTGCGG1GLASAI1DI1BET1AOERQGGDPK		
	PVA1IAHPDI1RQ1QEAR1RTE1RLAAT1ELATVMD1LAMEPEKE1RGD1QALSRWL		
	LP1VKN1GADAGFTVSNQAO1Q1IGAG1YREW1PE1QY1LDR1ARVWT1YEGT1GQAO1DL		
	LTRL1MEGGKGLDV1FLK1RARS1ELADYRS1QTRAA1AS1DLAFAES1RMA1LENEPET		
	GL1QANN1LY1AAM1EAP1AAM1TV1VEG1AD1QART1GAA"		
	complement(2914..3687)		
gene	/gene="AGR_C_2613"		
	complement(2914..3687)		
CDS	/gene="AGR_C_2613p"		
	/note="probable short-chain dehydrogenase PA2554 (imported) - Pseudomonas aeruginosa (strain PAO1)"		
	/codon_start=1		
	/transl_table=11		
	/product="AGR_C_2613p"		
	/protein_id="AAK87207.1"		
	/db_xref="GI:15156487"		
	/translation="MN1EGALVTGAASGLGAAVVA1MLA1RGA1V1TP1DNGE1AGKK		
	LA1B1IGKA1VQ1GDTSD1AD1QAO1A1IKVA1SKG1R1ILVNC1AG1GTAG1R1IGR1GQPL		
	GD1EQV1RW1L1GT1FN1MRL1AA1HMER1EQD1SD1QND1GV1VNT1AS1VAF1GQ1GQ		
	AA1YAS1KSG1GV1SL1P1AR1ELAR1FR1RVNT1VA1G1FL1PL1Q1G1PQ1E1VBS1LAG1Q1PH		
	PS1RG1DPA1E1AD1YV1R1FL1END1YNN1G1EV1R1DGA1IR1MQP"		
	complement(3684..5459)		
gene	/gene="AGR_C_2614"		
	complement(3684..5459)		
CDS	/gene="AGR_C_2614p"		
	/note="farnoloyl-CoA synthetase"		
	/codon_start=1		
	/transl_table=11		
	/product="AGR_C_2614p"		
	/protein_id="AAK87208.1"		
	/db_xref="GI:15156488"		
	/translation="M1WRNDPLG1PYPQK1NER1LLH1MCR1SAP1ERT1MADR1QGR1E1PWR1RV		
	SYA1E1DK1R1GQ1FL1D1HD1SV1ER1PL1V1SENS1E1HAL1VLA1O1HV1G1AS1A1IT1PAY		
	ATS1AD1LTK1AD1R1GQ1TR1PGM1VPA1ED1AT1PP1RA1GEV1DD1CT1PL1G1BN1P1ED1SN1TH		
	PE1TL1ET1EP1E1		

AAEENRARAIVIFGHDFHCAGLDLGEHIIHKTPIEGVRGSRMRHVAEFSQIEHGTIPW
 SALHGVVSGGLAEIATSTHLPVADBSAPFALPEGCGIIFVGGSGSVRIARILGVAMS
 DMILTRVAVSSAEERWNLVOYVPRGALAKOELATTAAASVAPISNVAVINALPRI
 QDMAKEDGLFVSEFISSTPTAASPEAEERLRAFLERKATTLKADTTA"

gene
 CDS
 complement (6511..7506)
 /gene="AGR_C_2619"
 complement (6511..7506)
 /gene="AGR_C_2619"
 /note=" (A011317) Metf protein"

/codon_start=1
 /transl_table=11
 /product="AGR_C_2619p"
 /protein_id="AK87211.1"
 /db_xref="GI:15156491"
 /translation="MYFRKILIEVAVSCILHNOVLSREKAMGMSFMDNSGADGVYRS
 SNILAGSTLEWMPRTAKIDDPNLLPDGRVYIAHIDGPRIADMTSKRLHDEFP
 VPHILPARGISGKNELEDMKRYOGAQTQALLGSGQKOPGSSSIDMETELF
 DKMGFTLHLAGHPEGNRDIDSDSSRGVQALLMODPASRTDARLALTOPFAEK
 VVLEMAKRIGNMGVITPVHVGAGPILQTLKFAIACGVSPILKYLQKRALDLSLF
 VYEPDFIAELIDHKAANDSLIEHVHIFPLGGIKPSABMKKHAFTGLANVA"

gene
 CDS
 7739..8473
 /gene="AGR_C_2620"
 /gene="AGR_C_2620"
 /note="probable transcription regulator PA4906 (imported)
 - Pseudomonas aeruginosa (strain PAO1)"

/codon_start=1
 /transl_table=11
 /product="AGR_C_2620p"
 /protein_id="AK87212.1"
 /db_xref="GI:15156492"
 /translation="MRQVDAATHGRAVIELREKILSGELPGSMRLEFVSTADLNDIS
 RFPVRALSRILEBGLNLPGGGFYRRPGFADVDVDAIEVRGMEGTARLAEBFP
 SKALREIATVQQLDCGDRVDYDFPGYALNRIPIHQALALGSEMIIRREVERA
 SSTLPFASPFAPDKANIGAFRRSLGAEQHRATVAAIAREGAREAVARHSRTA
 RTMLEYMRRAPELIAQVPLALISD"

gene
 CDS
 8623..10044
 /gene="AGR_C_2622"
 /gene="AGR_C_2622"
 /note="PROBABLE AMINOMETHYLTRANSFERASE (GLYCINE CLEAVAGE
 SYSTEM T PROTEIN)"

/codon_start=1
 /transl_table=11
 /product="AGR_C_2622p"
 /protein_id="AK87213.1"
 /db_xref="GI:15156493"
 /translation="MTMANSLSLEVLKQYPNPYEMLRNSKIGNTYVYPVAGERSNRS
 EQAMREGAVLPDOSHMDLIVEGPDAEKFLAHGINSFNTDLNRAKHVYTPDG
 HVIQDHIIFRERDKEFLVGRAPTSMLMCAVYKMYRIKIDPPSPRSEGRVLR
 THYROIQGPDAKFIPEKNGGPIPIKFPYVDMINIGSKVQALHSHSGAPLEVM
 GPVEDSYTILSCLEAKQAGVNLVCGSRAYSTNTPESMIPSPILGYTGGMKID
 YRDLGADSYEATGALGSPVSENIEDYVNPPELDYFTIGMKUDFYGKSLAMK
 GKTHRRKVTPEMNAEDVDVIAAFRPGEDHYKMDIPQANVASTGADILKGGKHVG
 MSFTGYSYNERCIIISLGIYDNDIQIGDVLTLVWGPBGDSAKSTERHKQKEIRVRV
 SEFPVAAERDNDVDSMRTKAS"

ORIGIN

Alignment Scores:

Pred. No.: 4.56e-82 Length: 10173
 Score: 1294.00 Matches: 266
 Percent Similarity: 61.19% Conservative: 84
 Best Local Similarity: 46.50% Mismatches: 220
 Query Match: 41.99% Indels: 2
 DB: 1 Gaps: 2

US-09-750-986D-30 (1-589) x AEO08067 (1-10173)

QY 5 GIUALLeuLeuProPheProGlyArgGILeueGluArgLeuGluHISThrPalalysThr 24
 Db 5444 GATCCGCTGGCCCTTATCGGAGAAATGAAATGAGGCTCTTCACTGCTGCTGG 5385
 QY 25 ArgProGluGluInThrCysValAlaIaArgAlaIaIaansGlyGluTrpArgArgIleSer 44

Db 5384 GCACCCGAGCGCACATGATGCGCCGACCTTCAGGGTCGAGCGCAAGCGCGCGGCTCAC 5325
 QY 45 TTTAAGluMeCpHeHisasnValArgAlaIaIaAGluSerLeuLeuProTyGlyLeu 64
 Db 5324 TACGCCAGGCTCTCGACAAATACGCGCTATCGCCATTCCTTTCATCATGACCTT 5265
 QY 65 SerAlaGluArgProLeuLeuIleValSerGlyAsnAspLeuGluHisLeuGluIleuAla 84
 Db 5264 TCGGTGAGAGGCGCGTGTGCTGCTTCTTGAATAATCCATGAGCATCGGTGATGGTG 5205
 QY 85 PheGlyAlaMetTyraIaGlyIleProTyCysProValSerProAlaTySerLeuLeu 104
 Db 5204 CTTCGCGCCACAGATGTCGATTCGCTCGGCTGCGCATCACCGCGCTTATGCC--ACG 5148
 QY 105 SerGluAspLeuAlaIleLeuAsnArgHisIleValGlyLeuLeuGluProGlyLeuValPhe 124
 Db 5147 TCGGCCGATTCACCAATCTCGCCGACATCCGCGGACAGATCACCCGCGGATGGTTT 5088
 QY 125 AlaAlaAspAlaIaProPheGluArgAlaIaIeGluThrIleLeuProAspAspValPro 144
 Db 5087 GCCGAGATGCGACGCGCTTTCGCGCGGCTCGGAGAGTTTTCAGACGATACACCG 5028
 QY 145 AlaIlePheThrArgGlyGlyLeuAlaGlyArg--ArgThrValSerPheAspSerLeu 163
 Db 5027 CTGCTCGCGCTCGCAACTGCGCGGAAAGATCGTTCCAATACGTTTCAATTCCGAAACCTG 4968
 QY 164 LeuGluGluProGlyGlyIleGluAlaAspAsnAlaPheAlaIaIaThrIleProAspThr 183
 Db 4967 CTTGAACCGAGCGCACCGACCGGTTGATCGGCTTTTGATGCGGTGCGGCGGATACC 4908
 QY 184 IleAlaIlePheLeuPheThrSerGlySerThrIleLeuProIleValAlaIaProThrThr 203
 Db 4907 GTCCGCAATTCCTTTCATCGTACGACGACGATCGTTCGCCAAGCGGTATCCAGAG 4848
 QY 204 GlnArgMetLeuCysAlaIaansGluIleMetLeuLeuGluInThrPheProValPheGlyGlu 223
 Db 4847 CAACGATGTTGTCCTCAATCAGAAATGATGCGATGTTATGCTTATCCGCGAG 4788
 QY 224 GluProProValLeuValAspTrpLeuProTrpAsnHisThrPheGlyGlySerHisAsn 243
 Db 4787 GAGCCGCGCGCTGCTTATATGGCGCCATGGACCAACAGCGCGCGGGAACAAGGTG 4728
 QY 244 IleGlyIleValLeuIleZasnGlyGlyThrTyTyTyLeuAspAspGlyIleProThrAla 263
 Db 4727 TTCAACCTGCTCTCTATATGGCGGTACTTATATGATGCGCGCAAGCCACACCG 4668
 QY 264 GlnGlyPheAlaGluThrLeuArgAsnLeuSerGluIleSerProThrAlaTyLeuThr 283
 Db 4667 GCGCAGATGCGCCAGACATGCAATCTCCGGAATATCGCCACTGCTATTTCAAC 4608
 QY 284 ValProIleGlyTyTrpGluIleValIaIaIeGluIaArgSerThrLeuArgIle 303
 Db 4607 GTTCCGCGTGGCCACAAATGCTGTGTCAGGACATCGCAAGGACGAGCGCTGCCCC 4548
 QY 304 ArgPhePheAlaArgMetIleLeuPhePhePheAlaIaIaAGlyLeuSerGlnGlyIle 323
 Db 4547 AGCTTTTTCGCGACCTGAAGATGCTATGCTCCGCGTCCGCGATCGCGACACACA 4488
 QY 324 TrpAspArgLeuAspArgValaIaGluGlnHisCysGlyGluArgIleArgMetMetAla 343
 Db 4487 TGGGATGCGCTGACGAGCTTTCATGACGACAGGTGCGCATGCGGTCTCATAGGTCGC 4428
 QY 344 GlyLeuGlyMetThrGluThrAlaProSerCysThrPheThrThrGlyProLeuSerMet 363
 Db 4427 GGCCTCGGCTCGACGAGACGACCCCTTCGCTGTTCGACCGCAACCGACGAGCAAG 4368
 QY 364 AlaGlyTyTrIleGlyLeuProAlaProGlyCysGluValIleLeuValProValaAspGly 383
 Db 4367 CCGGCAATATGGATGATCCCGGCGAGGCGGAGCATGAACCTGCGCTTCATGATGGC 4308
 QY 384 LysLeuGluGlyArgPheHisGlyProHisValaMetSerGlyTyTrpArgAlaProGlu 403
 Db 403


```

Db      4307 CGGTACCAATTACGCGTGAAGGCCCGCAACATCACGCCCGGTACTGCGCTAATGCGCAA 4248
Qy      404  GlnAsnAlaGlnAlaPheAspGluGluGlyTyrTyrCysSerSerGlyAspAlaIleLeuLeu 423
Db      4247 TTGACGGCGACGCGCTTCGACGAGAAAGGTTTCATGAATCGCGTACCGTCAAAATTT 4188
Qy      424  AlaAspProAlaAspProGlnIleGlyLeuMetPheAspGlyArgIleAlaGluAspPhe 443
Db      4187 GCGCGCGGACGACATCGCGCGCGCGCTTTATTTCATTCATGCGCCCATCGCGGAAACTTC 4128
Qy      444  LysLeuSerSerGlyValPheValSerValGlyProLeuArgThrArgAlaValLeuGlu 463
Db      4127 AACGTGCAGACCGCGACCTGGGTGCGCGCGCTTTGCGCGCGCATTAAGTCAACATG 4068
Qy      464  GlyGlySerTyrValLeuAspValValValAlaAlaPheAspArgGluCysLeuGlyLeu 483
Db      4067 TTGCGCGGCTGATTCCTGTGATGCCGTATACGGGCGAAGAACCGGCGCAACTGGCGCT 4008
Qy      484  LeuValPheProArgLeuLeuAspCysArgAlaLeuSerGlyLeuGlyValAlaSer 503
Db      4007 CTGCGTGTGCATTATCATACCGGCGCTACGTGAGCTTTGCGCGGTTCGCAACATCTTCC 3948
Qy      504  AspAlaGlyValLeuAlaSerGluProValArgAlaTrpPheAlaAspTrpLeuLysArg 523
Db      3947 GATGCGGAGATCATCCGACCCGAGTGTCAAGGCGCATGTCGCGAAGCTTTCGCGC 3888
Qy      524  LeuAsnArgGluAlaThrArgAlaSerArgIleMetTrpValGlyLeuLeuAspThr 543
Db      3887 CATCAAAAGACGCGAGCGGTTCGCAAGCGCGCTCATGCGCATTCGCGATGGAAGAC 3828
Qy      544  ProProSerIleAspIleGlyValThrAspIleGlySerIleAsnArgAlaVal 563
Db      3827 GCGCTCGCGCTCGAAGAAAGCGAGGTTACGACACAGAGATCATCAACGCGCGCTG 3768
Qy      564  LeuGlnTrpArgSerAlaLysValAspAlaLeuTyr 575
Db      3767 CTTCGTGCATCGCAAGAGCTGTAGAGTCACTTAC 3732

RESULT 15
AE009102/c 10173 bp DNA linear BCT 28-MAY-2004
LOCUS      Agrobacterium tumefaciens str. C58 circular chromosome, section 128
DEFINITION
ACCESSION  AE009102 AE008688
KEYWORDS   AE009102.1 GI:17739830
SOURCE     Agrobacterium tumefaciens str. C58
ORGANISM   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
            Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
REFERENCE  1 (bases 1 to 10173)
AUTHORS    Wood,D.W., Setubal,J.C., Kaul,R., Monks,D., Chen,L., Wood,G.E.,
            Zhou,Y., Woo,L., KitaJima,J.P., Okura,V.K., Almeida Jr.,N.F.,
            Zhou,Y., Bovee Sr.,D., Chapman,P., Cleddening,J., Deatherage,G.,
            Gillet,W., Grant,C., Guenther,D., Kutyavina,T., Levy,R., Li,M.,
            McClelland,B., Palmieri,A., Raymond,C., Rouse,G.,
            Saenphimmachak,C., Wu,Z., Gordon,D., Bisen,J.A., Paulsen,I.,
            Karp,P., Romero,P., Zhang,S., Yoo,H., Tao,Y., Biddle,P., Jung,M.,
            Krespan,W., Perry,M., Gordon-Kamm,B., Liao,L., Kim,S., Hendrick,C.,
            Zhao,Z., DoJan,M., Tingey,S.V., Tomb,J., Gordon,M.P., Olson,M.V.,
            and Nester,E.W.
            Submitted (27-SEP-2001) Department of Microbiology, University of
            Washington, 1959 NE Pacific Ave, Box 357242, Seattle, WA
            98195-7242, USA

```

```

FEATURES
    source
        1..10173
            /organism="Agrobacterium tumefaciens str. C58"
            /mol_type="genomic DNA"
            /strain="C58"
            /isolate="U. Washington"
            /db_xref="taxon:176299"
            /complement(100..1263)
            /gene="Atu1413"
            /note="Identified by sequence similarity; putative; ORF
            located using Blastx/Glimmer"
            /codon_start=1
            /transl_table=11
            /product="ABC transporter, substrate binding protein
            {branched chain amino acid}"
            /protein_id="AA142419.1"
            /db_xref="GI:17739831"
            /translation="MGCKTMNKRINTWMAVAVALTTAGSAPADVTIGTPTTGGCA
            ALGIPKQISEFWPADIDDKLVILLDAGDSVATYNNARRPASDKVDVIGSALT
            PAIAVAGVANETEPHFPALSPIPPTATKMFVMPQFSTLVKSMFARMQKGVKXV
            AFIGFADSGQGVQVDALKATGKDYGLVLEERYADTSVSGVLSKLAAPDAFI
            GASGTGALPQIGVRRERGRKGVVSGAVTFEFLIAGKAVNTVTPASGVNVAEDQ
            PDSAPRKQSGALTKAPBEKRGENTTQGSNHWMDVYLKRVPAALKAKKCTAEF
            RAALRDALSTGSEFNATHGVYEFADNHNHGLDQGVLLTAADGKTVLLPDNK"
            /complement(1317..2879)
            /gene="acd"
            /note="synonym: Atu1414"
            /complement(1317..2879)
            /gene="acd"
            /note="Identified by sequence similarity; putative; ORF
            located using Blastx/Glimmer"
            /codon_start=1
            /transl_table=11
            /product="acyl-CoA dehydrogenase"
            /protein_id="AA142420.1"
            /db_xref="GI:17739832"
            /translation="MELKATSGMRRIQSLRGDCDDALVVALINEARPAETVLEPIN
            TPADREGCTVVDGRVPTGFPADAPFRHAGDGLANDVBERFGGALPLVLQAAAPL
            FERGCVALLMAAGSTPAACHLAETAEAGVADDEVPRLLAAGEWATICTISEPPAGSDL
            ARLRSVAVREGMWRIEGRKIWISFGDHMMARIGICLARGTEETGTGLSLFVPD
            EIDGRNGTVTVERIEKMGHSGPTCALRFDQARGILLBEGRGALQUPARITELMLQ
            TGQGGILASALIDIAETVAQERQCGDPAKPAVAIAHHPDIRQJQEARLETTLA
            ATLELTAVNDLAMEPEKEBLRGDQLSLMLPLVYNFGADGFTVSNQAIQITGGAG
            YTEWEPLELDLAMEPKEBLRGDQLSLMLPLVYNFGADGFTVSNQAIQITGGAG
            RSKTKTAAASVIDAPAEBSRRMAALENEBETGLQNNYNNLNAWBAFAAMWTIVREGV
            ADAQARRTRKAA"
            /complement(2914..3687)
            /gene="Atu1415"
            /complement(2914..3687)
            /gene="Atu1415"
            /note="Identified by sequence similarity; putative; ORF
            located using Blastx/Glimmer"
            /codon_start=1
            /transl_table=11
            /product="3-hydroxyacyl-CoA dehydrogenase type II"
            /protein_id="AA142421.1"
            /db_xref="GI:17739833"
            /translation="NWIEGAGALVNGASGLCAAVVBMALARGAATTFIDRNGEAGKX
            LAAETGKAVQGDVTSDDADQAALIKYAAKAGKRLIUNCAIGTARGRLIGREGPPL
            GPBQVITRNLTGTFWMRLLAAHVAERQDSDRQDNGVTVNTASVAAPBEGQIGQ
            AAYVAKGQIVSALTAPAEALFRIRIVTAPGFLPTPLQGLPQEVESLAGQIPH
            PSRLGDPAEFADTVRELIENDYMGVIRLDGAIRMQPR"
            /complement(3684..5519)

```

gene

gene
CDS
/gene="Atu1416"
complement (3684. . 5519)
/gene="Atu1416"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon start=1
/transl_table=11
/product="long-chain-fatty-acid-CoA-ligase"
/protein_id="AA14242.1"
/db_xref="GI:17739834"
/translation="MSVLTMSPDVAMKEKPGNGEIMWRNDLGPYPOKLEIRLLHMC
RSPETWMAADRCGRPMRVRVYAAALDKLRIGQPLDHDLSVRPLIPLSENSIEHA
LMVLAQVIGLASAIIPTAYATSADTLKADIRGQITPGVPAEDATPPRALGEVD
DGTPLVGRANLPEDESRNTFHFETLLETPEAVADAPADVAPDTVAFLFTSGTGP
KAVITQSMCSQNMADVCYGFPEEPVVADAPMNHDTAGKRVNLVLYNGITY
IDRGPRSAQIQGLDNLNLDISPTWYFVPGAGHMLVOAMKDCALCRSFRDKMLM
YAGAAQMHOTDMLTELMTATVGHAVLMGAGLSTETAPSELFTCEPDQKNGIGPA
OCVTKVLPFGCRVELRLKGPNTIPGVRNELTAAAPDEGVRIGDTYKPAVDPP
RGGFPTDGRMAENFKLQGTWVAAGPLAQLVNNPAGLIRDAVITGMRBELGLVYP
FLPALRELVRSQHLSDAEILRHPSVRQIVAKLSAQKQKSSGASVRLIWMEDAL
REKEGVNDKSNORAVLHRLKEVLESTYGDTPVITGVREVA"
5664. . 6461
/gene="fad"
/note="synonym: Atu1417"
5664. . 6461
/gene="fad"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon start=1
/transl_table=11
/product="enoyl-CoA hydratase"
/protein_id="AA14242.1"
/db_xref="GI:17739835"
/translation="MSEKTEEPFVSYELRGDIALVGLNRPQKRAISDRFEVAILAAA
ARAENEAVALFGCHGRFCAGLIDGHIHTPTIEGRVGSRRMAVRSQIEHGIPIV
SALHNAVVGGLIELAASHTLRAVDEBSAFLPBEQRGI PVGGGSAVIALILVAARS
DMVLGVAASSAEAKRMNLVQYAVPRGEALAKQDELATIASNPVLSNVAINALPRT
OMAKEDGILFVPSFISFTPAASPEABERLRAFLERKATRLKAPDTTA"
complement (6511. . 7431)
/gene="mezf"
/note="synonym: Atu1418"
complement (6511. . 7431)
/gene="mezf"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon start=1
/transl_table=11
/product="methylentetrahydrofolate reductase"
/protein_id="AA14242.1"
/db_xref="GI:17739836"
/translation="WGMSSFWDNSGADGDVVRSSNIIAGSIEVMPRTAKIDFENI
LDPGRVYIAHIDGTPIADVSTAKRLHDEGF PVMPLPARGISGKNELDMLKRYOG
EAGYQALLIGDGKOPAKSPSSSIDMIETGLPKMGPETHLHAGPEGNRDIDSQS
SRGVQALLMKODFASRTDARLALTTPQPAFAKVIEMAKRIGMGTTIVVHGIAOP
TKLQTLIFALACGVGSLKVLQKRALDSSLFVYEPPTDFIAELIDHKANPDSILE
HVIHPILOGIKSAPAMKKAHSTGLANVA"
7739. . 8473
/gene="Atu1419"
7739. . 8473
/gene="Atu1419"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon start=1
/transl_table=11
/product="transcriptional regulator, Gntc family"
/protein_id="AA14242.1"
/db_xref="GI:17739837"
/translation="MRQVDAATHGRGAVITELREKILSGELPGQRMLFEVSTAELDIS
RIPVAESRLREGLNRLPGGFGVVRPFQAVVVAILEVGMGTARLAAERVA
SKVALEEDATVQGLDLCGRGRVDVDDVGAALNRI FHQIALCGSEWIRREVER
SSLPPASBSATLPDPRANIGAFRRSRKQOEQKAIVAIYARBARAEAVAREHSTR
RNLLEYMRREAPLEIAVPGIALISD"
8623. . 10044

CDS

```

/gene="gcvT"
/note="synonym: Atu1420"
8623..10044
/gene="gcvT"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=1
/product="glycine cleavage system T protein,
aminomethyltransferase"
/protein_id="AA14246.1"
/db_xref="GI:17739838"
/translation="MTYANSLETVLKQYENPVEMLNSTKGMVYVPVAGEPSNRS
EQAMRGVALFDSSHMDLIVGEPDAEKFLAHGINSFNPLNRKHPVPTPDC
HVGIDHIFERODKFLVGRAPSSWLMFCQAKXNVRILKDPSPSPREGVRVLR
THYVYQIQGDPARKIFEKNGGPIPDJLKFVWDVINIGSKVQALRHGMSAPGILEVY
GYVYDSKYLISLCLEAKQAGVNVLRGSGRAYSNTELESGLIPSLPQIYTGDMGLK
YPMIDGADSYETAGATGGSFVSENIEDYGVNPELIDYDFIYGMKKDPEIGKSALEMY
GKTRKKVTFEEMNAEDVVDVYASAFRGEEDHKRIDPQANYASTGADILKSGKNG
MSMETYSYNERCLSLGIVNDIQIDVLTLVGSEPDGSAKSTENHKKERVR
SETPYAARADNVADSWRTKAS"

Alignment Scores:
Score: 4.56e-82 Length: 10173
Percent Similarity: 1294.00 Matches: 266
Best Local Similarity: 61.19% Conservative: 84
Query Match: 46.50% Mismatch: 220
DB: 41.99% Indels: 2
Gaps: 2

US-09-750-986D-30 (1-589) x AE009102 (1-10173)
QY 5 GluAlaLeuLeuProPheProGlyArgIleLeuGluArgLeuGluNIstPAlaIbArThr 24
Db 5444 GATCGCGTGGCCCCCTATCCGCAAAATGTAATGAGCGTCTCTTCACTGTCGTCGTTGG 5385
QY 25 ArgProGluGlnThrCySvAlaAlaAlaArgAlaAlaAsnGlyGluTPRArgIleSer 44
Db 5384 GCACCGGAGGCGCATGATGATGCGCGACCGTCAGGCGTCGCGAGCCATGGCGCGGTCAGC 5325
QY 45 TyrAlaGluMetPheHisAsnValArgAlaIleAlaGlnSerLeuLeuProTyrGlyLeu 64
Db 5324 TACCCCGAGGCTCTCGACAAAGATACGGCGATACGGCCAGTTCCTTCTCGATCATGACCTT 5265
QY 65 SerAlaGluArgProLeuLeuIleValSerGlyAsnAspLeuGluNIstLeuGluAla 84
Db 5264 TCGGTCCAGAGGCGCGCTGCTGCTCTTCTGAATAATTCATCGACGATCGCTGATGCTG 5205
QY 85 PheGlyAlaMetYrAlaGlyIleProTyrCySvProValSerProAlaTyrSerLeuLeu 104
Db 5204 CTTCCGCGCCCGACATGTCGATATGCTCGGCTCGCATACGCGCGCTTATGCC--ACG 5148
QY 105 SerGlnAspLeuAlaYsleuArgHisGleValGlyLeuLeuGlnProGlyLeuValPhe 124
Db 5147 TCGGCCCATCTCACAAACTCGCCGACATCGCGGAGACATCACCCCGGGAGTGTTTT 5086
QY 125 AlaAlaAspAlaAlaProPheGlnArgAlaIleGluThrIleLeuProAspAspValPro 144
Db 5087 GCCAGAGATGAGAGCCCTTTCCGCGGGCGGTCCGCGAGAGTTTTCAGACGATGACACG 5028
QY 145 AlaIlePheThrArgGlyGluLeuAlaGlyArg--ArgThrValSerPheAspSerLeu 163
Db 5027 CTGCTCGCGCTCCGCAACTGCGCGAGAGATGTTCCAAATAGTTTCAATTTTGAAACCTTG 4968
QY 164 LeuGluGlnProGlyGlyIleGluAlaAspAsnAlaPheAlaAlaThrGlyProAspThr 183
Db 4967 CTTGAAACGAGCCGACCGAAGCGGTTGATCGTCTTTCATGTCGTCGCGCGGATAC 4908
QY 184 IleAlaYsPheLeuPheThrSerGlySerThrYsleuProYsAlaValProThrThr 203
Db 4907 GTGCCCAATTCCTTTTTCAGTCGAGGACGACATGCTTCCGCGCAAGCGGTATCCAGACG 4848
QY 204 GlnArgMetLeuCySvAlaAsnGlnIleMetLeuLeuGlnThrPheProValPheGlyGlu 223

```

```
Db 4847 CAAAGCATGTTGGTCCATTCAGAAATGATTCGATTTGATGCTACTTTCCGAG 4788
Qy 224 GTPProProValLeuValAaPTPLeuProTPAsnH1sThrPheGlySerHisAsn 243
Db 4787 GAACCCGCGGTCCGTGGATTGGCCCATGGAACACACGCGCGGGAACAGGTG 4728
Qy 244 ILGGLYLeValLeuTYrAsnGlyTYrThrTYrLeuAspAspGlyLysProThrAla 263
Db 4727 TTCAACTCGTCCCTATTAATGGGGTAACTATTAATCAATCGCGGCAAGCCGCGC 4668
Qy 264 GInGlyPheAlaGluThrLeuArgAsnLeuSerGluLeuSerProThrAlaTYrLeuThr 283
Db 4667 GCGCAAGATCGGCGCAGACATCTCGGAGACATATCGCCACCTGGTATTTCAAC 4608
Qy 284 ValProLysGlyTYrGlyGlyLeuValGlyAlaLeuGluAaPheSerThrLeuArgGlu 303
Db 4607 GTTCCCGCTGGCAAGAAATGCTGGTCCAGGCAATCGCAAGAGAGCGCGCTGCGCGC 4548
Qy 304 ArgPhePheAlaArgMetLysLeuPhePhePheAlaAlaGlyLeuSerGlnGlyIle 323
Db 4547 AGCTTTTTCGCACTGAAGATGCTGATGATGCTGCGGTGCGGATGCGCAGACACA 4488
Qy 324 TrpAspArgLeuAspArgValAlaGluGlnHisCysGlyGluArgIleArgMetMetAla 343
Db 4487 TGGGATGCGCTGACGAGGCTTTCATGCGCAACGCTCGCCATGCGCTTCATGCTGCC 4428
Qy 344 GlyLeuGlyMetThrGlyThrAlaProSerCysThrPheThrThrGlyProLeuSerMet 363
Db 4427 GGCCTCGGCTCGACGAGAGCGCACCTTCTCGCTGTTGACCGAAACCGCAGACAG 4368
Qy 364 AlaGlyTYrTLeGlyLeuProAlaProGlyCysGluValLysLeuValProValAspGly 383
Db 4367 CCGGCAATTCGATATCCGCGCAGGCGCTGACGATGAACCTGCGCCCTTGATGGC 4308
Qy 384 LysLeuGluGlyArgPheHisGlyProHisValMetSerGlyTYrTrpArgAlaProGly 403
Db 4307 CGGTAGGAATTAAGCGCGAAGCGCGAAGCATCAGCCCGTTACTGGCGTAATGGCGAA 4248
Qy 404 GlnAsnAlaGlnAlaPheAspGlyGlyTYrTYrCysSerGlyAspAlaIleLysLeu 423
Db 4247 TTGACGCGCGCACCTTCGACGAGGAGGTTCTATCGAAATCGGCGATACCGTCAATTT 4188
Qy 424 AlaAspProAlaAspProGlnLysGlyLeuMetPheAspGlyArgIleAlaGluAspPhe 443
Db 4187 GCCGTGCGGACGATCCGCGCGCTTTTATTTCATGCGCGCATGCGGAAACCTTC 4128
Qy 444 LysLeuSerSerGlyValPheValSerValGlyProLeuArgThrArgAlaValLeuGlu 463
Db 4127 AAGCTGCAAGACCGGACACTGGGTGGCGGTGGCGCTTGGCGCGCAATTAGTCAACATG 4068
Qy 464 GlyGlySerTYrValLeuAspValValAlaAlaPheAspArgGlyCysLeuGlyLeu 483
Db 4067 TTCGCGCGGCTGATCCGTATGCCGTATACGAGGCGAAACCGGCGCAACTGGCGCT 4008
Qy 484 LeuValPheProArgLeuLeuAspCysArgAlaLeuSerGlyLeuGlyLysGluAlaSer 503
Db 4007 CTGCTGCTGCATTCATACCGGCGCTACGTAGCTGTTCGCGGTTCGCAACATCTCTCC 3948
Qy 504 AspAlaGlyValLeuAlaSerGluProValArgAlaTrpPheAlaAspTrpLeuLysArg 523
Db 3947 GATCGGAGATCATCCGCGACCCGAGTGTCAAGGCGCAGATTCGCGAAAGCTTTCGCGC 3888
Qy 524 LeuAsnArgGluAlaThrGlyAsnAlaSerArgIleMetTrpValGlyLeuLeuAspThr 543
Db 3887 CATCAAAACAGCGCGGCTTCGCGCAAGCGCGTATGCGCATTCGCGATGGAAGAC 3828
Qy 544 ProProSerIleAspLysGlyGlyValThrAspLysGlySerIleAsnGlnArgAlaVal 563
Db 3827 GCGCTGCGCTTCGAGAAAGCGAGGTTACGACAAAGGATCGATCAACAGCGCGCGTG 3768
Qy 564 LeuGlnTrpArgSerAlaLysValAspAlaLeuTYr 575
```

Db 3767 CTTCGATCGCAAGAGCTGTAGATCACTTAC 3732

Search completed: February 8, 2005, 02:03:13
Job time : 6717 sec

This Page Blank (uspio)